# Do Chimpanzees Reason According to the Disjunctive Syllogism?

Jan M. Engelmann [a], Lou M. Haux [b], Christoph Völter [c], Hanna Schleihauf [a, d], Josep Call [e], Hannes Rakoczy [f]*, and Esther Herrmann [g]*

[a] Department of Psychology, University of California, Berkeley, 94705 Berkeley, CA, USA

[b] Center for Adaptive Rationality, Max Planck Institute for Human Development, 14195 Berlin, Germany

[c] Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, University of Vienna, 1210 Vienna, Austria

[d] Department for Primate Cognition, Georg-August-University Goettingen, 37077 Goettingen, Germany

[e] School of Psychology and Neuroscience, University of St. Andrews, KY16 9AJ, St. Andrews, UK

[f] Department of Developmental Psychology, Georg-Elias Müller Institute of Psychology, Georg-August-University Goettingen, 37073 Goettingen, Germany

[g] Department of Psychology, University of Portsmouth, PO1 2DY, Portsmouth, UK

* denotes shared last authorship

Correspondence:
Jan Engelmann
2121 Berkeley Way
University of California, Berkeley, 94705 Berkeley, CA
jan_engelmann@berkeley.edu

## ABSTRACT

Psychologists disagree about the development of logical concepts such as *or* and *not*. While some theorists argue that infants reason logically, others maintain that logical inference is contingent on linguistic abilities and emerges around age 4. In this Registered Report, we conducted five experiments on logical reasoning in chimpanzees. Subjects (N=16; ten females; M=24 years) participated in the same setup that has been administered to children: the two-, three-, and four-cup-task. Chimpanzees performed above chance in the two-cup-, but not in the three-cup-task. Furthermore, chimpanzees selected the logically correct option more often in the test than the control condition of the four-cup-task. We discuss possible interpretations of these findings and conclude that our results are most consistent with non-deductive accounts.

In a famous story by the Stoic logician Chrysippus, a dog pursuing a rabbit arrives at a fork with three paths. The dog fails to track a scent on the first path, so moves to sniff the middle path, also fails to track a scent, following which she chases immediately down path C, without sniffing. Chrysippus wondered: is the dog engaging in a logical inference - *a or b or c, not a or b, therefore c* – or is she using a simpler cognitive strategy?

Chrysippus eventually endorsed the second option. This view is shared by many modern theorists, who maintain that the ability to engage in logical inference distinguishes the thought of adult humans from the thought of nonhuman animals and prelinguistic infants (Bermúdez, 2007; Floridi, 1997; Oelze, 2018). Given that most models of logical reasoning rely on logical concepts expressed via linguistic terms, the core concepts of classical logic (such as *or* and *not*) are argued to be beyond the representational abilities of nonverbal organisms (Schechter, 2013; Seitz, 2020).

Recently, the question of the relationship between linguistic ability and logical thought has attracted new interest due to reports suggesting that a preverbal population can reason according to the disjunctive syllogism. Infants as young as 12 months were shown to potentially engage in logical inference (Cesana-Arlotti et al., 2018, see also 2020): When two objects (a dinosaur and a flower) were hidden in different locations, and one location was shown to contain one of the objects (the dinosaur) infants looked longer – indicating surprise – when the second location was subsequently revealed to contain the same object, rather than the other object (the flower). One interpretation is that infants generate this prediction by disjunctive syllogism: *a or b, not a, therefore b*. Alternative interpretations of the data, however, suggest that infants might be following non-deductive strategies rather than making a logical inference. Specifically, infants might run a single simulation of which object is hidden in a given location and revise if necessary (Leahy & Carey, 2020), or engage in object tracking (Jasbi et al., 2019).

These more parsimonious explanations for the infant data gain support from other experimental paradigms which indicate that it is not until children are more linguistically competent, at around age 4, that they can reason according to the disjunctive syllogism (Leahy & Carey, 2020; Mody & Carey, 2016). Whether, and if so, to what extent, nonverbal organisms engage in logical reasoning thus remains an open question.

Here we contribute to this discussion by studying logical thought in one of our closest living relatives, chimpanzees. Three sources of evidence suggest that chimpanzees might engage in logical inference (Schloegl & Fischer, 2017; Völter & Call, 2017). First, experimental paradigms in which chimpanzees can draw on evidence to infer what likely happened provide evidence for diagnostic inference (*if a then b, b therefore a*). In a study by Völter & Call (2014), for example, chimpanzees spontaneously used the trail a piece of food had left behind – the food's 'traces' – to draw conclusions about its current location. However, although sometimes described as a logical inference, the antecedent does not follow logically from the consequent in abductive reasoning, but rather involves an inference to the best explanation (Sober, 2013). What is inferred is possibly, but not necessarily true and therefore does not have the same validity as logical principles. Second, stronger evidence that chimpanzees reason logically comes from studies of tool selection (Tomasello, 2014; Völter & Call, 2017) in which subjects infer according to a form of modus ponens (*if a then b, a therefore b*). When presented with a number of different tools which vary in terms of key properties, chimpanzees reliably and flexibly select the tool that is most appropriate to the task at hand – even when the problem to be solved is in a different room, out of sight (Manrique et al., 2010). One interpretation of this finding is that chimpanzees make a predictive inference based on modus ponens reasoning: if I possess the appropriate tool, then I will obtain the food. Third, one of the best pieces of evidence for logical inference comes from Call's cup task (Call, 2004, 2006; Hill et al., 2011), which suggests that chimpanzees reason in a manner that is consistent with disjunctive syllogism (*a or b, not a, therefore b*). In this experimental paradigm, an experimenter hides a piece of food in one of two opaque cups. Then, during the demonstration phase, they present the subject with visual evidence about where the reward is not hidden: they lift one of the cups and reveal it to be empty. Subjects' behavior in the choice phase is highly consistent. In nearly 100% of trials, chimpanzees select the other cup (Call, 2004, 2006). This pattern of behavior might be indicative of logical inference: subjects produce a new mental representation (the food is in B) on the basis of the combination of two previously held representations (the food is either in A or in B and the food is not in A).

Similarly to the infant data discussed above, however, chimpanzees' performance in the cup task is consistent with other, non-deductive mechanisms, which vary significantly in their cognitive demands. Following Mody and Carey (2016), two such mechanisms can be distinguished: 'avoid empty' and 'maybe A, maybe B'. According to the first alternative interpretation, chimpanzees merely avoid the empty cup. Like many other mammals, chimpanzees might follow a heuristic of continuing to forage when they do not encounter food in a given location and thus select the other cup more or less accidentally, as it were, and without representing the fact that it must contain the reward. But this does not look like a serious rival hypothesis to the disjunctive syllogism interpretation. For chimpanzees make the relevant inference with regard to the food's location also when they first observe how two different types of food are hidden in two locations (apple in cup A and banana in cup B) and are then shown a piece of food that used to be in one of the two cups (e.g. the apple). Subjects in this setup – comparable to the infant study reviewed above – never see the empty cup, but still reliably choose the correct location (i.e. cup B; Call, 2006; Premack & Premack, 1994)[1]. The ‘maybe A, maybe B’ account poses a more serious alternative. It is predicated on the notion that chimpanzees represent two possible locations of the food but not their dependent relationship. When one of the locations is shown to be empty (“not A”), subjects are left with “maybe B” and so go for the second cup. Cognitively speaking, this analysis makes fewer demands on the reasoning subject than the logical account in that it does not involve the representation of a relationship between the two possible locations of food (seeing that A is not the case does not affect the inferred probability that B is the case). In addition, it does not involve the generation of a new representation: subjects select the other cup because it might contain the reward (“maybe B”), and not because it – as the logical inference would have it – *necessarily* contains it. Based on existing evidence, it is not possible to rule out that chimpanzees solve the cup task by reasoning according to the “maybe A, maybe B” mechanism.

Luckily, however, the development of an experimental extension of the cup task in children has provided us with exactly the right tool to determine whether chimpanzees in fact solve the cup task by reasoning according to the disjunctive syllogism (Mody & Carey, 2016). The main methodological innovation is to present participants with twice the

[1] Note that “avoid empty” might also be conceptualized in terms of avoiding a location that is only represented to be empty (based on inferential reasoning), but never actually seen as empty. In this case, the alternative account cannot be ruled out by prior research.

number of options: Two pairs of two cups (the so-called four-cup task). Participants are shown, during the demonstration phase, that one reward is hidden in one pair of cups (A, B) and one reward is hidden in a second pair of cups (C, D). Then, one of the cups (A) is revealed to be empty. The disjunctive syllogism and the "maybe A, maybe B" hypotheses make contrasting predictions. A logically reasoning agent infers that B *must* contain the reward and so chooses this option; an agent reasoning according to the simpler alternative chooses B, C, or D with equal probability. 3-, 4-, and 5-year-old children show the former pattern. 2.5-year-old children, in contrast, show the latter (although children at this age choose B significantly more often than expected by chance in the two-cup task). This result is important because it shows that it is possible, in practice, to display competent performance in the original two-cup task without representing the disjunction between A and B.

Gautam, Suddendorf, and Redshaw (2021), however, argue that successful performance in the original four-cup task is not sufficient to demonstrate logical reasoning. Notice that one potential alternative interpretation of positive results in the four-cup task is in terms of local enhancement. By highlighting that cup B is empty, the experimenter might draw subjects' attention to the first pair of cups, inadvertently increasing the likelihood that subjects choose cup A next. In order to rule out this low-level explanation, Gautam and colleagues (2021) introduce the *reveal baited cup* version of the four-cup task. Participants are shown, just like in the classic version of the four-cup task, that one reward is hidden in one pair of cups (A, B) and one reward is hidden in a second pair of cups (C, D). Then, in contrast to the classic version, one of the cups (A) is revealed to be baited and the reward is discarded. A logically reasoning agent – but not an agent who is influenced by local enhancement – will consequently choose C or D with equal probability. The new *reveal baited cup* version of the four-cup task thus helps to rule out the local enhancement alternative interpretation. Importantly, there is empirical evidence that it is possible to pass one version of the four-cup task but not the other. As Gautam, Suddendorf, and Redshaw (2021) report, 2.5-, 3-, 4-, and 5-year-old children perform competently in the *reveal empty cup* version, but only 5-year-old children additionally succeed at the *reveal baited cup* version.

To our knowledge, there is only one previous investigation of the four-cup task in nonhuman primates. Ferrigno, Huang, & Cantlon (2021) present evidence that three olive baboons succeed in the *reveal empty cup* version. The same monkeys, however, do not succeed in the *reveal baited cup* version, leaving open the 'stimulus enhancement' alternative interpretation discussed in the previous paragraph.

**The Current Experiments.**

In the current Registered Report, we investigated logical inference in chimpanzees (the preregistration can be found here: https://osf.io/4mxbd/). All reported methods and analyses were preregistered unless specified otherwise.

Subjects participated in five experiments: the two-cups task, the three-cups task, two versions of the four-cups task, and a follow-up study (see Figure 1). Experiment 1 is a replication of the two-cup task (Call, 2004, 2006). A reward is hidden in one of two cups (A, B), one cup is shown to be empty (A), and the question is whether chimpanzees pick the other cup (B) above chance (chance level = 0.5). Based on previous research, we predicted that chimpanzee will be at or near ceiling in their selection of cup B (Call, 2004).

As argued above, successful performance in the two-cup task is explainable in terms of a variety of underlying cognitive processes. We ran four further experiments to zero in on the mechanism used by chimpanzees.

Experiment 2 involves the three-cup task. In this task, subjects are presented with three cups (A, B, C) and two items of food. One item of food is hidden in cup A and the other item is hidden in either B or C. The question of interest is whether chimpanzees are above chance in their selection of the option that *must* contain the food (A), relative to the options that *could* contain a reward (B, C). In determining baseline or chance levels against which to compare performance, we followed recent suggestions, made on theoretical grounds, by Leahy and Carey (2020). The most basic, baseline possibility for choosing non-logically is random selection of one the three possible cups (chance level would be set at 33%). But a theoretically more relevant way of choosing non-logically is to select either side with a probability of 0.5 (for details on this account, see the discussion). Thus, in line with Leahy and Carey's (2020) proposal to analyze children's and non-human primates' performance in the three-cup and related tasks with this baseline possibility as the relevant reference value, we set the chance level at 50%.

In Experiments 3 and 4, chimpanzees were exposed to the two versions of the four-cup task (see Figure 2). In both versions, we compared chimpanzees' behavior in a test condition to a control condition. Half of the subjects started with Experiment 3: the *reveal empty cup* version (Mody & Carey, 2016). In the test condition, one item of food is hidden in a first pair of cups (cup A or B) and a second food item is hidden in a second pair of cups (cup C or D). One of the four cups is then revealed to be empty (B). If chimpanzees reason according to the disjunctive syllogism, they should selectively choose the other cup of the same pair (A). In the control condition, again two food items are hidden in the four cups but without any visible cup pairings (so that subjects only know that two items are hidden in A, B, C or D). Like in the test condition, one cup (B; yoked to test condition) is then revealed to be empty. Subjects can then only infer that two items are hidden in A, C, or D.

The other half of subjects started with Experiment 4: the *reveal baited cup* version (Gautam, Suddendorf, Redshaw, 2021). The test condition is identical to the *reveal empty cup* version, except that one cup is revealed to be baited (B) and the associated reward is discarded. If chimpanzees reason according to the disjunctive syllogism, they should choose cup C or D with equal probability (since they can infer that cup A must be empty). In the control condition, again two food items are hidden in the four cups but without any visible cup pairings (so that subjects only know that two items are hidden in A, B, C or D). Like in the test condition, one cup (B; yoked to test condition) is then be revealed to be baited and the reward is discarded. Subjects can thus only infer that there is one item left in A, C, or D.

Experiment 5 was a follow-up study (preregistered, but not part of the original Registered Report). In an experimental setup with reduced task demands, we directly compared chimpanzees' responses in the *reveal empty cup* version to their responses in the *reveal baited cup* version (see Ferrigno, Huang, & Cantlon, 2021).

# EXPERIMENT 1

## *Methods*

**Participants.** Sixteen chimpanzees (ten females), living at Ngamba Island Chimpanzee Sanctuary, Uganda, ranging in age from 12 to 31 years, M = 24 years participated in Experiment 1. Chimpanzees have access to a large outdoor enclosure during the day and receive regular daily feedings, daily enrichment, and water *ad libitum*. Subjects voluntarily participated in the study and were never deprived of food or water. For more information on subjects, please refer to table S1 of the supplementary material. Testing for Experiments 1-5 took place between June and August 2021.

**Materials.** Testing took place in two adjacent rooms: the observation room and the choice room. The rooms were connected by a door, which could be opened or closed. Two cups were positioned outside of the choice room (see Figure 1a). The cups were placed at a distance of 210cm from each other[2] and at a distance of 100cm from the choice room. Each cup was connected to a rope, which extended into the choice room. Chimpanzees could access a cup and its content by pulling the appropriate rope. Half an apple was used as reward. During the observation, a black occluder (240cm x 50cm x 50cm) was used to conceal the baiting process.

[2] In Experiment 1, the two cups that form one assortment are placed at a distance of 210cm from one another, while they are placed at a distance of 70cm in Experiment 3. This is done in order to ensure that subjects don't learn a simple rule in Experiment 1 ("always pick the cup right next to the empty cup") and then apply this rule in Experiment 3.

**Procedure.** Each trial consisted of two phases, an observation phase and a choice phase. During the observation phase, the subject was located in the observation room. The experimenter (E) started the trial by placing one piece of apple on the ground in front of the subject (but outside of the subject's reach). Next, E lifted and turned upside down the two cups to demonstrate to the subject that they were empty. E proceeded to cover the two cups with an occluder, thereby preventing the subject from observing the hiding process. E picked up the piece of apple and baited one of the cups in the following way. She first held the apple above the center of the occluder, calling the subject's name while doing so. She grabbed the apple with both hands, lowered her hands, and, once her hands were hidden behind the occluder, moved to one of the cups (keeping her hands behind the occluder) and placed the piece of apple under the cup. Then she moved to the second cup (again keeping her hands behind the occluder) and also lifted and manipulated the second cup (so that subjects could not infer where the apple was hidden). Whether E baited the second or the first cup was counterbalanced across trials. E now showed her empty hands to the subject. E then removed the occluder. Once E had removed the occluder, she demonstrated to the subject that one of the cups was empty by opening it and showing the inside of the cup (which cup was empty was counterbalanced across trials), before placing it back in its original position. In order to avoid stimulus enhancement, E also touched the other cup. We counterbalance across trials whether E touched the empty or the baited cup first. Finally, a second experimenter (E2) opened the door connecting the observation room and the choice room (it took subjects approximately three seconds to move between rooms). This represented the end of the observation phase.

The choice phase started once subjects moved from the observation room to the choice room. Crucially, when subjects entered the choice room they were automatically centered such that they were equidistant between the two cups. In the choice room, subjects were able to access the contents of one cup. Once the subjects had made a choice by pulling one of the ropes, E removed the remaining rope.

Subjects participated in a total of twelve trials, distributed across two sessions (6 trials per session). Each session took place on a different day.

**Inclusion Criteria and Coding.** Once chimpanzees made a choice by selecting one of the cups, this choice was coded. If chimpanzees, for whatever reason, did not make a choice within 30 seconds of opening the door, the trial was repeated. This happened on three trials for one chimpanzee. If chimpanzees did not make a choice on three consecutive trials, the session was stopped and the missing trials were repeated on the next day (this never happened). As mentioned above, chimpanzees participated in two sessions of 6 trials. If chimpanzees did not reach the final trial number of 12 within six sessions, data collection for this chimpanzee was stopped (this never happened).

Whether chimpanzees selected the cup which necessarily contained an apple was coded live by the first experimenter. A research assistant, unaware of the study design and hypothesis, independently coded 25% of all trials from video. Interrater agreement was perfect (Cohen's $\kappa = 1$).

----------Figure 1 goes here-------------

### *Results*

To test whether chimpanzees chose the correct cup above chance in the two–cup task, we compared the *choice of the target cup* to the hypothetical chance level of 0.5 by fitting an intercept-only model, with *subject ID* as a random intercept and *trial* (z-transformed) in *subject ID* as a random slope (including the correlations between random slopes and intercept). Note that analyses presented in this manuscript represent a confirmatory effort.

Chimpanzees performed significantly above chance in the the two–cup task (intercept-only GLMM Estimate ± SE: 4.234 ± 1.058, $z = 4.003$, $p < 0.001$, see Table A1 in SI). More specifically, chimpanzees chose the correct cup in 95% of trials (for individual performance see Figure 3A).

## EXPERIMENT 2

### *Methods*

**Participants.** Those subjects that selected the baited cup significantly above chance in Experiment 1 (two-tailed binomial test: $p<0.05$) participated in Experiment 2. Since two chimpanzees did not fulfill this criterion, the sample size for Experiment 2 was 14 chimpanzees.

**Materials.** The setup of Experiment 2 was very similar to the setup of Experiment 1. The main difference was that Experiment 2 involved three cups (see Figure 1b). One cup stood on its own (single cup location), while the two other cups formed an assortment (two-cup location). The single cup was positioned 140cm from the two-cup location and the cups within the two-cup location were placed at a distance of 70cm from each other. Whether the single cup was located on the left or the right (from the perspective of the observation room) was counterbalanced across trials. The cups were again placed 100cm from the observation room. Each cup was connected to a rope, which extended into the choice room. Chimpanzees could access a cup and its content by pulling the appropriate rope. Half an apple was used as reward. During the observation, two black occluders (100cm x 50cm x 50cm) were used to conceal the baiting process.

**Procedure.** Each trial consisted of two phases, an observation phase and a choice phase. During the observation phase, the subject was located in the observation room. The experimenter (E) started the trial by placing two pieces of apple on the ground. Next, E lifted and turned upside down the three cups to demonstrate to the subject that they were empty. E proceeded to cover the single-cup location and the two-cup location with separate occluders, thereby preventing the subject from observing the hiding process. E first placed one piece of apple in the cup at the single-cup location and then the other piece of apple in one of the two cups at the two-cup location (order of baiting and choice of baited cup at the two-cup location were counterbalanced across trials). E baited the cup at the two-cup location in the following way. She first held the apple above the center of the occluder, calling the subject's name while doing so. She grabbed the apple with both hands, lowered her hands, and, once her hands were hidden behind the occluder, separated them and moved each hand to one cup (so that subjects could not see where the apple was hidden). She showed her empty hands to the subject. E picked up the second apple and repeated the exact same sequence of behaviors to bait the cup at the single-cup location. E then removed both occluders. Finally, E2 opened the door connecting the observation room and the choice room and stepped to the side. This represented the end of the observation phase.

The choice phase started once subjects moved from the observation room to the choice room. In the choice room, the subjects were able to access the contents of one cup. Once the subject had made a choice by pulling one of the ropes, the experimenter and a second experimenter removed the two remaining ropes.

Subjects participated in a total of twelve trials, distributed across two sessions (6 trials per session). Each session took place on a different day.

**Inclusion Criteria and Coding.** Once chimpanzees made a choice by selecting one of the cups, this choice was coded. If chimpanzees, for whatever reason, did make a choice within 30 seconds of opening the door, the trial was repeated (this never happened in Experiment 2). If chimpanzees did not make a choice on three consecutive trials, the session was stopped and the missing trials were repeated on the next day (again, this never happened in Experiment 2). As mentioned above, chimpanzees participated in two sessions of 6 trials.

Whether chimpanzees selected the single cup which necessarily contained an apple or one of the cups at the two-cup location that could contain a piece of apple was coded live by the first experimenter. A research assistant, unaware of the study design and hypothesis, independently coded 25% of all trials from video. Interrater agreement was perfect (Cohen's $\kappa = 1$).

## ***Results***

To test whether chimpanzees chose the correct cup above chance in the three–cup task, we compared the *choice of the target cup* to the hypothetical chance level of 0.5 by fitting an intercept-only model, with *subject ID* as a random intercept and *trial* (z-transformed) in subject ID as a random slope (including the correlations between random slopes and intercept).

Chimpanzees did not perform significantly above chance in the three–cup task (intercept-only GLMM Estimate ± SE: 0.048 ± 0.155, $z$ = 0.309, $p$ = 0.757, see Table A2 in the SI). More specifically, chimpanzees chose the correct cup in 51% of trials (for individual performance, see Figure 3A). No individual performed significantly ($p<0.05$) above chance according to a two-tailed binomial test.

----------Figure 3 goes here-------------

## EXPERIMENT 3

### *Methods*

**Participants.** The same 14 chimpanzees who participated in Experiment 2 participated in Experiment 3. However, one chimpanzee stopped participating. Thus the sample size for Experiment 3 was 13 chimpanzees. To account for potential order effects, 6 chimpanzees, upon completion of Experiment 2, continued with Experiment 4 and then participated in Experiment 3. Seven chimpanzees started with Experiment 3 and then participated in Experiment 4.

**Materials.** The same materials as in Experiment 2 were used in Experiment 3. The only difference was that there were two two-cup locations (and therefore a total of four cups). See Figure 1c.

**Procedure.** "*Reveal empty cup*" version: Chimpanzees participated in a test and a control condition. The general procedure of the test condition was identical to the procedure of Experiment 1 (except for there being four cups in Experiment 3). E first took one piece of food and hid it in one of the two first cups. E then took a second piece of food and hid it in one of the two last cups. Once E had baited both assortments, she demonstrated to the subject that one of the four cups was empty by turning it upside down, shaking it, and showing the inside of the cup, before placing it back in its original position. In order to avoid stimulus enhancement, E also touched the three remaining cups (we counterbalanced the order in which E touched the cups). In the test condition, subject can thus infer that one of the cups must contain a reward and that the two other cups might contain a reward. The control condition was identical to the test condition except that the four cups formed one group, and not two, as in the test condition. E first hid two pieces of food, one after the other, and then revealed the empty content of one of the cups. In the control condition, subjects can thus only infer that there are two food items hidden in three possible cups. Subjects should pick randomly between the three cups.

In a within-subjects design, subjects participated in a total of twelve trials in each condition, distributed across four sessions (6 trials per session). Half of the subjects first participated in the test condition and then in the control condition (AABB) and the other half of subjects followed the reverse pattern (BBAA). Each session took place on a different day.

**Inclusion Criteria and Coding.** Once chimpanzees made a choice by selecting one of the cups, this choice was coded. If chimpanzees, for whatever reason, did not make a choice within 30 seconds of opening the door, the trial was repeated. This occurred for one out of the 14 chimpanzees, who stopped participating from their first trial of Experiment 3 onwards. Data collection for this chimpanzee was stopped.

In the test condition, whether chimpanzees selected the target cup (the cup next to the cup which was revealed to be empty) was coded live by the first experimenter. In the control condition, the first experimenter also coded whether chimpanzees select the target cup (this was yoked to the test condition: for each trial, the target cup in the control condition was the same cup that was the target cup in the corresponding trial of the test condition). A research assistant, unaware of the study design and hypothesis, independently coded 25% of all trials from video. Interrater agreement was perfect (Cohen's $\kappa = 1$).

----------Figure 2 goes here-------------

## ***Results***

To investigate chimpanzees' choice of the correct cup in the *reveal empty cup* task, we compared subjects' *choice of the target cup* in the test condition to that in the control condition. We formulated a full model with the predictors *condition* (test, control), *age* (in years), *sex* (female, male), *trial number* within *condition* and *order of condition* (coded as factor: control-first, test-first) as fixed effects and *subject ID* as a random intercept. As random slopes, we included *condition* and *trial number* within *subject ID* (including the correlations between random slopes and intercept). The covariates *age* and *trial number* were z-transformed and *condition* was treatment-coded (with the control condition as reference category).

The full model fit the data significantly better than the null model which lacked the effects of *condition, age* and *sex* ($\chi^2 = 8.552$, $p = 0.036$, see table A3 in the SI). *Condition* ($\chi^2 = 8.544$, $p = 0.003$) had a significant effect on performance, suggesting that chimpanzees chose the correct cup more often in the test compared to the control condition, see Figure 3B. More specifically subjects chose the correct cup in 48% of trials in the test condition and in 29% of trials in the control condition. Additionally, *order of condition* ($\chi^2 = 4.434$, $p = 0.035$) had a significant effect on the performance, suggesting that chimpanzees chose the correct cup more often when the control condition was presented first. There was no effect of *age* ($\chi^2 = 0.013$, $p = 0.908$), *sex* ($\chi^2 = 0.010$, $p = 0.920$), nor *trial* ($\chi^2 = 1.659$, $p = 0.198$).

## EXPERIMENT 4

### *Methods*

**Participants.** The same chimpanzees that participated in Experiment 3 participated in Experiment 4 (N=13).

**Materials.** The same materials as in Experiment 3 were used in Experiment 4.

**Procedure.** '*Reveal baited cup*' version: Chimpanzees participated in a test and a control condition. The procedure of the test condition was identical to the procedure of the '*reveal empty cup*' version except that a baited cup was uncovered. E first took one piece of food and hid it in one of the two first cups (A or B). E then took a second piece of food and hid it in one of the two last cups (C or D). Once E has baited both pairs, she removed the piece of food from one of the baited cups. E saliently lifted the cup, took the food, placed it in a nearby container (out of the chimpanzee's reach), and, finally, placed the cup back in its original position. In order to avoid stimulus enhancement, E also touched the three remaining cups (we counterbalanced the order in which E touched the cups). The control condition was identical to the test condition except that the four cups formed one assortment, and not two, as in the test condition. E first hid two pieces of food, one after the other, and then removed the food from one of the cups. Here, chimpanzees had a $^{2}/_{3}$ chance of choosing one of the target cups (the two cups of the pair which was still baited).

In a within-subjects design, subjects participated in a total of twelve trials in each condition, distributed across four sessions (6 trials per session). Half of the subjects first participated in the test condition and then in the control condition (AABB) and the other half of subjects followed the reverse pattern (BBAA). Each session took place on a different day.

**Inclusion Criteria and Coding.** Once chimpanzees had made a choice by selecting one of the cups, this choice was coded. No trials were repeated or excluded.

In the test condition, whether chimpanzees selected one of the two target cups (the cups which represented the other pair, next to the cup from which the food was removed) was coded live by the first experimenter. In the control condition, the first experimenter also coded whether chimpanzees selected one of the target cups (this was yoked to the test condition: for each trial, the target cups in the control condition were the same cups that were the target cups in the corresponding trial of the test condition). A research assistant, unaware of the study design and hypothesis, independently coded 25% of all trials from video. Interrater agreement was perfect (Cohen's κ = 1).

## *Results*

To investigate chimpanzees' choice of the correct pair in the *reveal baited cup* task, we compared subjects' *choice of the other pair*, i.e. the cup pair from which food had not been removed, in the test condition to that in the control condition. We formulated a full model with the predictors *condition* (test, control), *age* (in years), *sex* (female, male), *trial number within condition* and *order of condition* (coded as factor: control-first, test-first) as fixed effects and *subject ID* as a random intercept. As random slopes, we included *condition* and *trial number* within *subject ID* (including the correlations between random slopes and intercept). The covariates *age* and *trial number* were z-transformed and *condition* was treatment-coded (with the control condition as reference category).

The full model fit the data better than the null model which lacked the effect of *condition, age* and *sex* ($\chi^2 = 14.933$, $p = 0.002$, see table A3). *Condition* ($\chi^2 = 3.957$, $p = 0.047$) had a significant effect on performance, suggesting that chimpanzees chose the other pair more often in the test compared to the control condition, see Figure 3C. More specifically subjects chose the other pair in 85% of trials in the test condition and in 75%

of trials in the control condition. Additionally, older chimpanzees were significantly more likely to choose the other pair ($\chi^2$ = 6.447, $p$ = 0.011). There was no effect of *sex* ($\chi^2$ = 0.033, $p$ = 0.855), *trial number* ($\chi^2$ = 0.935, $p$ = 0.334), nor of *order of condition* ($\chi^2$ = 1.039, $p$ = 0.308).

*Comparison of the two test conditions.* As a secondary analysis, we fit another binomial GLMM to compare the performance in the test conditions of Experiment 3 and 4. The dependent variable for this analysis was chimpanzees' *choice of the other pair*, i.e. the cups next to the cup which was shown to be empty (Exp. 3: reveal-empty) or from which the food was removed (Exp. 4: reveal-baited). As predictor variables, we included *test conditions* (reveal-empty, reveal-baited), *age*, *sex*, *trial number within condition*, the *order of experiments* (coded as factor: Exp3-first, Exp4-first) and *subject ID* as a random intercept. As random slopes, we included *test condition* and *trial number* within *subject ID* (including the correlations between random slopes and intercept). The covariates *age* and *trial* number were z-transformed and *test condition* was treatment-coded (with the reveal baited condition as reference category).

The full model fit the data better than the null model which lacked the effect of *test condition, age* and *sex* ($\chi^2$ = 15.867, $p$ = 0.001, see table A5 in the SI). *Test condition* ($\chi^2$ = 10.134, $p$ = 0.001) had a significant effect on performance, suggesting that chimpanzees chose the other pair significantly more often in the reveal baited compared to the the reveal empty cup task, see Figure 3D. Additionally, older chimpanzees were significantly more likely to select the other pair ($\chi^2$ = 5.724, $p$ = 0.017). There was no effect of *sex* ($\chi^2$ = 0.608, $p$ = 0.435), *trial* ($\chi^2$ = 0.771, $p$ = 0.380), nor of *order of experiment* ($\chi^2$ = 0.204, $p$ = 0.651).

## EXPERIMENT 5

Note that Experiment 5 was originally not part of this Registered Report. The main goal of Experiment 5 (preregistered on the Open Science Framework: https://osf.io/6pg5z/?view_only=c1c5b32c05944ba3a790d4267a1bcedd) was to investigate whether chimpanzees would perform better in the *reveal empty* version of the four-cup task in a setup with reduced task demands. We directly compared *reveal empty*, as the test condition, to *reveal baited*, as the control condition (see Ferrigno, Huang, & Cantlon, 2021). To reduce working memory demands, the experimenter, upon revealing that one of the cups was empty (or baited in the control condition), left the cup in the open position (i.e., did not close the cup again, as in Experiment 3). Chimpanzees thus had a constant visual aid reminding them which cup did not contain the reward. In addition, we also placed the four cups on one table and closer to each other (compared to Experiment 3).

### *Methods*

**Participants.** Eight chimpanzees participated in Experiment 5. Four of the chimpanzees had already participated in Experiment 1-4. The other four chimpanzees were exposed to the four-cup task for the first time. We did not detect any difference in performance between experienced and naïve subjects (see Results). We had a within-subjects design. Chimpanzees were exposed, in counterbalanced order, to each condition (*reveal empty* and *reveal baited*).

**Materials.** Testing took place in one room. Four cups were positioned outside of the room on a table. The two cups that formed a pair were placed at a distance of 15cm from each other. The two pairs were placed at a distance of 30cm from each other. The backside of the cups was removed, so that the experimenter could place the food rewards inside the cups without having to move the cups.

**Procedure.** *Reveal empty*: At the beginning of the procedure, the four cups were placed on the table with the backside of the cups facing the chimpanzees (so that chimpanzees could look inside the cups and see that they were empty). Two pieces of apple were also on the table. The experimenter (E) called the chimpanzee and turned the four cups around (so that chimpanzees could not look inside anymore). Then E took one of the pieces of apple, hid it inside her hand (which formed a fist), first moved her hand into one cup, remained in the cup for two seconds, removed her hand from the cup, showed the closed hand to the chimpanzee, and then moved her hand into the second cup, again remained in the cup for two seconds, took her hand out of the cup and revealed to the chimpanzee that her hand was empty. Whether E placed the food in the first or second cup was counterbalanced across trials and subjects. Then E took the second piece of apple and repeated the procedure, hiding the food in one of the two cups that formed the second pair.

Next, E turned around an empty cup (which cup was turned around was counterbalanced across trials and subjects) such that the open backside was facing the chimpanzee. Finally, E pushed the table towards the subject. Once the subject had made a choice by pointing at one of the cups, the experimenter turned around that cup, handed the chimpanzee the piece of apple (if the subject had picked a cup with food), and then pulled the table back again. E removed all remaining food from the cups and placed them again in the initial position (open backside facing subject) before starting the next trial.

*Reveal baited*: The procedure in *reveal baited* was identical to *reveal empty* except that E turned around a baited cup, took the apple that was placed inside it, and put the apple into a nearby food container.

Subjects participated in a total of sixteen trials in each condition, distributed across four sessions (8 trials per session). Each session took place on a different day.

**Inclusion Criteria and Coding.** Once chimpanzees made a choice by selecting one of the cups, this choice was coded. If chimpanzees, for whatever reason, did make a choice within 30 seconds of pushing the table towards them, the trial was repeated (this never happened in Experiment 5). If chimpanzees did not make a choice on three consecutive trials, the session was stopped and the missing trials were repeated on the next day (again, this never happened in Experiment 5).

Whether chimpanzees selected a cup of the other pair – the pair that was not manipulated by the experimenter – was coded live by the first experimenter. A research assistant, unaware of the study design and hypothesis, independently coded 25% of all trials from video. Interrater agreement was perfect (Cohen's κ = 1).

### *Results*

To investigate chimpanzees' choice of the other pair in the *reveal empty* and *reveal baited cup* task, we ran a GLMM with binomial error distribution and logit link function using the function *glmer* of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015).

We compared subjects' performance in the reveal-empty to that in the reveal-baited condition. We formulated a full model with the predictors *condition* (test: reveal empty, control: reveal baited), *age* (in years), *sex* (female, male), *trial number within condition* and *order of condition* (coded as factor: control-first, test-first) as fixed effects and *subject ID* as a random intercept. As random slopes, we included *condition* and *trial number* within *subject ID* (including the correlations between random slopes and intercept). The covariates *age* and *trial number* were z-transformed and *condition* was treatment-coded (with the control condition as reference category).

The full model fit the data significantly better than the null model which lacked the effect of *condition, age* and *sex* ($\chi^2 = 18.288$, $p < 0.001$, see figure 4 and table A6 in the SI). *Condition* ($\chi^2 = 15.988$, $p < 0.001$) had a significant effect on performance, suggesting that chimpanzees chose the other-pair more often in the reveal-baited compared to the reveal-empty condition, see Figure 4. More specifically subjects chose the other-pair in 86% of trials in the reveal-baited and in 52 % of trials in the reveal-empty condition. There was no effect of *age* ($\chi^2 = 1.648$, $p = 0.199$), *sex* ($\chi^2 = 0.875$, $p = 0.350$), *trial* ($\chi^2 = 0.006$, $p = 0.937$), nor of *order of condition* ($\chi^2 = 1.601$, $p = 0.206$).

----------Figure 4 goes here-------------

## DISCUSSION

Across five experiments, we investigated chimpanzees' ability to reason logically. Chimpanzees successively participated in the two-cup task, the three-cup task, and two versions of the four-cup task. In addition, in a follow-up experiment, we exposed chimpanzees to a version of the four-cup task with reduced working memory demands. In short, we found that chimpanzees performed significantly above chance (set at 50%) in the two-cup task; at chance (set at 50%) in the three-cup task; and significantly better in the test compared to the control conditions of the four-cup task (*reveal empty* and *reveal baited*). The subjects' performance was nearly identical in both versions of the four-cup task – the original (Experiments 3 & 4) and the follow-up with lowered task demands (Experiment 5).

The near-ceiling performance in the two-cup task (95% correct choice of the other cup) is in line with prior research (Völter & Call, 2017). As reviewed in the introduction, success in the two-cup task is compatible with a number of different underlying reasoning mechanisms. The finding that chimpanzees did not appreciate the fact that one cup in the three-cup task must, by logical necessity, contain a reward – as evidenced by their chance level performance (chance was set at 50%) – raises doubts about the possibility that chimpanzees solve the two-cup task by logical thought; it is also in line with prior research (Hanus & Call, 2014). We compared chimpanzees' choices in the three-cup task to a conservative hypothetical chance level of 50%, rather than the less demanding chance level of 33%. A comparison to the latter chance level would have resulted in a significant difference (see SI). Independent of the appropriate chance level in the three-cup task, however, a group-level average choice of 51% of the certain option doesn't provide strong evidence that chimpanzees infer that one of the three cups must contain a reward.

In the four-cup task, chimpanzees participated in a test and a control condition. Chimpanzees' performance in the four-cup task seems, at least at first sight, compelling support for logical processing: in both *reveal empty* and *reveal baited*, chimpanzees made the choice that is in line with logical inference significantly more often in the test compared to the control condition. The comparison to the control condition is crucial as it allows us to rule out low-level interpretations, for example that chimpanzees in the test condition of *reveal empty* simply picked the cup next to the one revealed to be empty. Importantly, these results also present clear evidence against the two other alternative interpretations of successful performance in the two-cup task discussed in the introduction, 'avoid empty' and 'maybe A, maybe B', which both predict that subjects pick any of the remaining cups with a probability of 33%.

It might seem that chimpanzees performed better, in absolute terms, in *reveal baited* (chimpanzees made the correct choice in 85% of trials) than in *reveal empty* (chimpanzees made the correct choice in 48% of trials). But it is important to point out that (1) the DV in *reveal baited* was different from the DV in *reveal empty* and (2) the difference between test and control condition is in fact larger in *reveal empty* (Experiment 3) than in *reveal baited* (Experiment 4). Yet, absolute performance is relevant to the interpretation of the current results, and it is noteworthy that chimpanzees' choice of the cup that by logical necessity must contain a reward consistently approximated 50%: chimpanzees chose the target cup in 51% of trials in Experiment 2, in 48% of trials in Experiment 3, and again in 48% in Experiment 5. What are we to make of this relatively low performance? Is it even correct to speak of low performance?

Ferrigno, Huang, and Cantlon (2021) ran a four-cup task with olive baboons and found similar performance to the one that we observed here. Yet, they concluded that baboons reason according to the disjunctive syllogism, whereas in our opinion the current results do not present strong evidence for logical reasoning in chimpanzees. What is going on? The key to understanding this discrepancy is that Ferrigno and colleagues base their conclusion on the comparison between the *reveal empty* and *reveal baited* conditions and a comparison to chance level, which was set at 33%. We on the other hand compare performance to a baseline level of 50%. We chose this baseline level because a comparison to 33% opens the door to three alternative interpretations of chimpanzee's performance that do not involve logical inference.

One is the '*minimal representation of possibility*' proposal (Leahy & Carey, 2020). This hypothesis, which was developed to account for the performance of 2- and 3-year-old children in the cup tasks and other related tasks, the Y-shaped tube task (Beck et al., 2006; Redshaw & Suddendorf, 2016; Robinson et al., 2006) and partial ignorance tasks (Kim et al., 2016; Kloo et al., 2017; Rohwer et al., 2012), maintains that children below the age of 4 do not represent possibilities as possibilities but as facts. Two- and 3-year-old children in the four-cup task, the argument goes, track that one reward is hidden in one pair of cups (A, B) and another reward is hidden in a second pair of cups (C, D). When they see, in *reveal empty*, that, for example, A does not contain a reward, they simply learn that A does not contain a reward, but nothing more. They now make two simulations: that B contains a reward and that C (or D in 50% of cases) contains a reward, treat these simulations as facts, and then randomly choose one of the two cups that they 'know' to contain a reward. Likewise, in *reveal baited*, those with minimal cognitive representation skills first track the two pieces of reward and subsequently track how one of the rewards (e.g. C) is removed. Then they guess, and treat as fact, that A (or B, in 50% of cases) contains the reward. Based on this reasoning, the '*minimal representation of possibility*' account predicts that agents with minimal representation choose (1) the certain cup with a probability of 50% in both the *three-cup* task and the *reveal empty* version of the four-cup task (because they believe they know which cup in each assortment contains the reward and then pick randomly) and (2) the other pair in reveal baited with a probability close to 100% (B. Leahy, personal communication). Thus, based on the present results, chimpanzees, like children below the age of 4, might only have a '*minimal representation of possibility*': they simulate which cups contain food and then treat that simulation as actual. However, although the current findings are in line with the minimal account, it is unclear whether this proposal can explain other evidence suggesting that chimpanzees act in such a way as to accommodate multiple possibilities (Engelmann et al., 2021) and that chimpanzees prefer a single baited cup to a set of six cups (one of them baited; see Hanus & Call, 2014).

A second possible account of chimpanzees' performance in the various versions of the cup tasks presented here proposes that subjects approach the task in terms of locations rather than individual cups. Consider the four-cup task. Chimpanzees might represent that food is *here* (in the pair of cups A and B) and that food is *there* (in the pair of cups C and D). In *reveal baited*, chimpanzees then see that the food from A is removed, and with it the thought 'food is *here*', leaving them with the single representation: 'food is *there*', and consequently pick either C or D. In *reveal empty*, chimpanzees observe that A is empty, so both representations are still in place – food being *here* and *there* – and so chimpanzees select either of the two locations randomly. The same rationale can explain chimpanzees' performance in the three-cup task. The advantage of this account is that it can explain the performance rate in both versions of the four-cup task and the three-cup task (it is also closely related to the minimal account described in the previous paragraph but does not involve a commitment to the idea that chimpanzees treat their guesses as facts). However, it is again unclear whether this perspective can explain chimpanzees' decisions in other, closely related tasks. Hanus and Call (2014), for example, found that chimpanzees follow a probability ratio and consider both the number of hidden rewards and the number of hiding locations when choosing between different assortments.

The third alternative interpretation is probabilistic updating (Hanus & Call, 2014; Rescorla, 2009). The probabilistic updating account places emphasis on the finding that chimpanzees perform better in the test compared to the control condition of *reveal empty*. This finding can be explained as follows. Chimpanzees might not represent a logical relationship between cup A and cup B, but a probabilistically dependent relationship. When chimpanzees see, in *reveal empty*, that one of the cups in one pair, say A, does not contain food, they update the probability that B contains food. This interpretation of chimpanzees' behavior is attractive because it strikes a middle ground: it is not as cognitively demanding as thought that employs logical operators and it is not as low-level as the alternative described in the previous paragraph. Yet, the probabilistic updating account also has one disadvantage relative to the '*minimal representation of possibility*' account: it doesn't predict the approximately 50% level of target cup choice that we observed in the three-cup task, the *reveal empty version* in Experiment 3 and the *reveal empty version* in Experiment 5. In fact, it is unclear what performance levels the probabilistic account would predict exactly in the current experiments. In addition, chimpanzees' performance in a metacognitive search task is not in line with probabilistic updating: when a reward is hidden in A, B, or C, and chimpanzees acquire information that the reward is not in A or B, they nevertheless search for more information before choosing C on most trials (Call & Carpenter, 2001).

One final option is that chimpanzees are in fact able to reason logically, but that various performance factors prevented them from demonstrating this ability in the three-cup task and *reveal empty*. As other authors have highlighted (e.g. Mody & Carey, 2016), the three- and four-cup task place high demands on participants in terms of working memory and attentional span. For example, even the simplified procedure of Experiment 5 requires subjects to pay uninterrupted attention to a complex series of events for approximately 20 seconds. Even short bouts of inattentiveness might cause subjects to miss key information (e.g. where a piece of food has been placed). While we cannot fully rule out this interpretation, one of our findings suggests that task demands are not the whole story: Chimpanzees showed identical absolute performance in a version of *reveal empty* with reduced task demands compared to a version of *reveal empty* with increased task demands (see Experiment 5 compared to the test condition of Experiment 3). In addition, there is strong evidence that chimpanzees' short term memory in similar experimental setups is excellent (Amici et al., 2010; Völter et al., 2019). Independent of these considerations, one key challenge for future research is to develop nonverbal tests of logical reasoning that require less advanced executive function skills.

Our experimental setup closely matches the setup used in previous studies with children, allowing us to compare the performance of chimpanzees to the performance of children at different ages. In the three-cup task, chimpanzees chose the certain cup on 51%

of trials, which is in-between the performance of 2.5- (47%) and 3-year-old children (60%), but note that children, in contrast to chimpanzees, received additional training with this task (Mody & Carey, 2016). In *reveal empty*, chimpanzees selected the target cup on 48% of trials. Three-year-olds did so on 58% of trials, 4-year-olds on 64% of trials, and 5-year-olds on 76% of trials in the study by Mody & Carey (2016) and, in the study by Gautam and colleagues (2021), 2.5-year-olds did so on 72% of trials, 3-year-olds on 76% of trials, 4-year-olds on 80% of trials and 5-year-olds on 82% of trials. In *reveal baited*, chimpanzees chose the other pair in 86% of trials, while 2-year-olds did so on 54% of trials, 3-year-olds on 60% of trials, 4-year-olds on 74% of trials, and 5-year-olds on 98% of trials (Gautam, Suddendorf, Redshaw, 2021). This comparison suggests that chimpanzee thought, at least as revealed by performance on the current tasks, is not clearly in line with that of either 2-, 3-, 4- or 5-year-old children. In the three-cup task and *reveal empty*, chimpanzees look like 3-year-old or younger children. In *reveal baited*, however, chimpanzees are more similar to 4- and 5-year-old children (Gautam, Suddendorf, & Redshaw, 2021).

To conclude, let's return to the question that motivated the current investigation. Do chimpanzees reason according to the disjunctive syllogism? The present results provide only weak evidence in support of this possibility. Especially the relatively low likelihood of picking the option that must, by logical necessity, contain a reward in the three-cup task and *reveal empty* of the four-cup task make this interpretation of the current results unlikely. Yet, nonetheless, the present findings allow us to rule out a number of alternative interpretations of successful performance in the cup task and simultaneously raise several interesting questions for future research. Chimpanzees' relatively poor performance – from an adult human perspective – in the three-cup task and *reveal empty* provide fruitful starting points for developing a theory of chimpanzee thought processes. As they stand, the results seem to provide empirical support for an intuition the Stoic logician Chrysippus had more than 2000 years ago: that nonhuman animals do not reason in line with the disjunctive syllogism.

**ACKNOWLEDGEMENTS**

We thank Joshua Rukundo and all the staff at Ngamba Island Chimpanzee Sanctuary: Innocent Ampaire, Phillip Sekulya, Dorothy Basemera, John Mutambo, Joseph Kaale, Joseph Masereka, Paul Neynje, Rashia Nansubuga, Stanley Hategeka, Stanley Kyama and Titus Mukungu. Special thanks go to Ben Apamuku for building the apparatus. Finally, we thank Anna-Claire Schneider for helpful comments on a previous version of this manuscript. The data and R scripts associated with this manuscript are available on GitHub: https://github.com/cvoelter/disjunctive_syllogism.

## REFERENCES

Amici, F., Aureli, F., & Call, J. (2010). Monkeys and apes: Are their cognitive skills really so different? *American Journal of Physical Anthropology*, *143*(2), 188–197. https://doi.org/10.1002/ajpa.21305

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's Thinking About Counterfactuals and Future Hypotheticals as Possibilities. *Child Development*, *77*(2), 413–426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Bermúdez, J. L. (2007). *Thinking without words* (1st paperb. ed). Oxford Univ. Press.

Call, J. (2004). Inferences About the Location of Food in the Great Apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus). *Journal of Comparative Psychology*, *118*(2), 232–241. https://doi.org/10.1037/0735-7036.118.2.232

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition*, *9*(4), 393–403. https://doi.org/10.1007/s10071-006-0037-4

Cesana-Arlotti, N., Kovács, Á. M., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature Communications*, *11*(1), 5999. https://doi.org/10.1038/s41467-020-19734-5

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263–1266. https://doi.org/10.1126/science.aao3539

Engelmann, J. M., Völter, C. J., O'Madagain, C., Proft, M., Haun, D. B. M., Rakoczy, H., & Herrmann, E. (2021). Chimpanzees consider alternative possibilities. *Current Biology*, *31*(20), R1377–R1378. https://doi.org/10.1016/j.cub.2021.09.012

Ferrigno, S., Huang, Y., & Cantlon, J. F. (2021). Reasoning Through the Disjunctive Syllogism in Monkeys. *Psychological Science*, *32*(2), 292–300. https://doi.org/10.1177/0956797620971653

Floridi, L. (1997). Scepticism and Animal Rationality: The Fortune of Chrysippus' Dog in the History of Western Thought. *Archiv Für Geschichte Der Philosophie*, *79*(1). https://doi.org/10.1515/agph.1997.79.1.27

Gautam, S., Suddendorf, T., & Redshaw, J. (2021). When can young children reason about an exclusive disjunction? A follow up to. *Cognition*, *207*, 104507. https://doi.org/10.1016/j.cognition.2020.104507

Hanus, D., & Call, J. (2014). When maths trumps logic: Probabilistic judgements in chimpanzees. *Biology Letters*, *10*(12), 20140892. https://doi.org/10.1098/rsbl.2014.0892

Hill, A., Collier-Baker, E., & Suddendorf, T. (2011). Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. *Journal of Comparative Psychology*, *125*(1), 91–103. https://doi.org/10.1037/a0020867

Jasbi, M., Bohn, M., Long, B., Fourtassi, A., Barner, D., & Frank, M. C. (2019). *Comment on Cesana-Arlotti et al. (2018)* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/g2h7m

Kim, S., Paulus, M., Sodian, B., & Proust, J. (2016). Young Children's Sensitivity to Their Own Ignorance in Informing Others. *PLOS ONE*, *11*(3), e0152595. https://doi.org/10.1371/journal.pone.0152595

Kloo, D., Rohwer, M., & Perner, J. (2017). Direct and indirect admission of ignorance by children. *Journal of Experimental Child Psychology*, *159*, 279–295. https://doi.org/10.1016/j.jecp.2017.02.014

Leahy, B. P., & Carey, S. E. (2020). The Acquisition of Modal Concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

Manrique, H. M., Gross, A. N.-M., & Call, J. (2010). Great apes select tools on the basis of their rigidity. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(4), 409–422. https://doi.org/10.1037/a0019296

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48. https://doi.org/10.1016/j.cognition.2016.05.012

Oelze, A. (2018). *Animal rationality: Later medieval theories, 1250-1350*. Brill.

Premack, D., & Premack, A. J. (1994). Levels of causal understanding in chimpanzees and children. *Cognition*, *50*(1–3), 347–362. https://doi.org/10.1016/0010-0277(94)90035-3

Redshaw, J., & Suddendorf, T. (2016). Children's and Apes' Preparatory Responses to Two Mutually Exclusive Possibilities. *Current Biology*, *26*(13), 1758–1762. https://doi.org/10.1016/j.cub.2016.04.062

Rescorla, M. (2009). Chrysippus' dog as a case study in non-linguistic cognition. In *The Philosophy of Animal Minds* (pp. 52–71). Cambridge University Press.

Robinson, E. J., Rowley, M. G., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's Sensitivity to Their Own Relative Ignorance: Handling of Possibilities

Under Epistemic and Physical Uncertainty. *Child Development*, *77*(6), 1642–1655. https://doi.org/10.1111/j.1467-8624.2006.00964.x

Rohwer, M., Kloo, D., & Perner, J. (2012). Escape From Metaignorance: How Children Develop an Understanding of Their Own Lack of Knowledge: **Overestimation of Own Knowledge in Young Children**. *Child Development*, *83*(6), 1869–1883. https://doi.org/10.1111/j.1467-8624.2012.01830.x

Schechter, J. (2013). Could Evolution Explain Our Reliability about Logic? In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology Volume 4* (pp. 214–239). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672707.003.0008

Schloegl, C., & Fischer, J. (2017). *Causal Reasoning in Non-Human Animals* (M. R. Waldmann, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.36

Seitz, F. (2020). A Mind Selected by Needs: Explaining Logical Animals by Evolution. *Acta Analytica*. https://doi.org/10.1007/s12136-020-00421-5

Sober, E. (2013). *Core questions in philosophy: A text with readings* (6th ed). Pearson Education.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

Völter, C. J., & Call, J. (2014). Great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo abelii) follow visual trails to locate hidden food. *Journal of Comparative Psychology*, *128*(2), 199–208. https://doi.org/10.1037/a0035434

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: Perception, learning, and cognition.* (pp. 643–671). American Psychological Association. https://doi.org/10.1037/0000012-029

Völter, C. J., Mundry, R., Call, J., & Seed, A. M. (2019). Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907), 20190715. https://doi.org/10.1098/rspb.2019.0715

# FIGURE LEGENDS

**Figure 1**. Schematic drawing of the experimental setup in Experiment 1, 2, 3 and 4. In Experiment 1, chimpanzees made a choice between two cups (see Picture a). In Experiment 2, chimpanzees were presented with three cups (see Picture b). Finally, in Experiments 3 and 4, chimpanzees were exposed to four cups (see Picture c).

**Figure 2**. Schematic drawing of the experimental setup in Experiment 3 (a) and Experiment 4 (b). Test conditions are depicted on the left, control conditions on the right. Notice that the difference between test and control conditions was that the four cups formed two assortments in test conditions, and one assortment in control conditions. The placement of the rewards was yoked across conditions (i.e., the same cups contained rewards across the two conditions).

**Figure 3.** Dot and box plot of the chimpanzees' performance in Experiment 1-4. The dots represent mean individual values. The error bars show the bootstrapped 95% CI of a GLMM with all predictor variables centered except for condition; the filled circle on the error bar shows the model prediction. The horizontal, dashed line represents the hypothetical chance level. A. Proportion of target cup choices in the two-cup (Experiment 1) and three-cup task (Experiment 2). B. Proportion of target cup choices in the *reveal empty cup* task (Experiment 3). C. Proportion of other pair choices in the *reveal baited cup* task (Experiment 4). D. Proportion of other pair choices in the test conditions of the *reveal empty cup* task (Experiment 3) and *reveal baited cup* task (Experiment 4).

**Figure 4.** Dot and box plot of the chimpanzees' other pair choices in Experiment 5. The dots represent mean individual values. The error bars show the bootstrapped 95% CI of a GLMM with all predictor variables centered except for condition; the filled circle on the error bar shows the model prediction.