Adults' and Children's Comprehension of Disjunction in a Guessing Game

Masoud Jasbi[1] & Michael C. Frank[1,2]

[1] Harvard University

[2] Stanford University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Masoud Jasbi, Postal address. E-mail: masoud_jasbi@fas.harvard.edu

Abstract

Enter abstract here. Each new line herein must be indented, like this line.

Keywords: conjunction, disjunction, implicatures, semantics, pragmatics, logical connectives, language, acquisition, development, children

Word count: X

Adults' and Children's Comprehension of Disjunction in a Guessing Game

Introduction

In this chapter, I examine the proposed differences between adults and children's interpretation of the disjunction word *or*. Previous research has suggested that adults and children might differ in their interpretation of *or* in two ways. First, unlike adults, children might interpret *or* as logical conjunction, akin to *and* (Singh, Wexler, Astle-Rahim, Kamawar, & Fox, 2016; Tieu et al., 2016). Second, children might interpret *or* as inclusive disjunction while adults interpret it as exclusive (Crain, 2012). In this chapter, I present three studies that assess adults and children's understanding of *and* and *or* in a guessing game paradigm. These studies show that four-year-olds' interpretation of conjunction and disjunction may not be as different from adults as previously supposed.

Study 1 tested adults' interpretations of logical connectives in the context of a guessing game using Two and Three-Alternative Forced Choice judgment tasks (2AFC and 3AFC). The results showed that adults interpret *and* and *or* differently. They interpreted *and* as conjunction and *or* as inclusive disjunction. However, in the task with three alternatives (3AFC) adults did not consider a disjunction felicitous when both disjuncts were true. Comparing the 2AFC and 3AFC results, we find that the felicity of disjunctive statements is sensitive to the measurement. 2AFC task systematically underestimated judgments of felicity and better approximated truth judgments compared to the 3AFC task. This finding is intuitive given that more options provide a better opportunity to express nuances of linguistic interpretation.

Study 2 investigated children's judgments in the same guessing game as study 1 using a 3AFC task. I used three alternatives to give children a better chance of expressing their pragmatic knowledge and judgments of felicity (Katsos & Bishop, 2011). The study also analyzed and categorized children's open-ended spontaneous feedback to the guesser. Both the 3AFC judgments and the categories of open-ended responses showed that four-year-olds differentiated *or* from *and*. While children's judgments in the 3AFC task showed no sign of

infelicity for disjunctive guesses when both disjuncts were true, their open-ended feedback showed that children find such guesses infelicitous. In their open-ended feedback, children's comments showed that use of a conjunction in such cases would be more appropriate.

Study 3 used the same paradigm as study 2, but focused on replicating children's open-ended responses and contrasting them with the results of a 2AFC task. As in study 2, both truth judgments and open-ended feedback showed that children differentiated *or* from *and*. The 2AFC task showed no evidence that children find disjunctions with true disjuncts infelicitous. However, children's judgments did not differ significantly from those of adults in the 2AFC task of study 1. As in study 2, children's open-ended feedback suggested that when both disjuncts are true, children find a disjunctive statement infelicitous and the conjunctive alternative more appropriate. Overall, the results of study 2 and 3 show that forced-choice judgement tasks underestimate children's pragmatic competence. Therefore, using open-ended elicitation and analysis of children's feedback **along with** forced choice judgment tasks may provide a better understanding of children's true semantic and pragmatic knowledge.

The studies reported here build on previous studies, and fill two gaps in the literature as well. First, most previous research focused on children's interpretation of *or* in complex sentences – for example with other logical words such as quantifiers *every* and *none*. Here, I test children and adults' understanding of *and* and *or* in simple existential sentences like "*There is a cat or a dog.*" To my knowledge, only Braine and Rumain (1981) used simple existential constructions before, but their experimental paradigm was relatively more complex than the paradigm used here. As discussed before, simplifying the paradigm is an important step in reducing conjunctive interpretations that arise due to non-linguistic strategies. Second, most previous research tested children and adults using 2AFC truth value judgment tasks (Crain & Thornton, 1998). Here, I report adults and children's judgments on both 2AFC and 3AFC tasks. I also use children's open-ended spontaneous feedback to develop relevant analytical response categories and I replicate the findings in a following

pre-registered study. Katsos & Bishop (2011) argued that 3AFC judgment tasks are better suited for assessing children's pragmatic competence. I present results that suggest even a 3AFC task can underestimate children's pragmatic knowledge and that children's spontaneous and open-ended elicited responses provide valuable insights not available in forced choice judgments.

Study 1: Adult Judgments

The goal of this study was to examine adults' interpretations of *and* and *or* as a benchmark for children's interpretations. I designed the study as a guessing game. Participants saw a card, read a description, and had to evaluate the description with respect to what they saw on the card. In test trials, the descriptions contained the conjunction word *and* and the disjunction word *or*. I tested adults in both two-alternative and three-alternative forced choice tasks (2AFC and 3AFC). The results suggested that adults interpreted *and* as a conjunction and *or* as an inclusive disjunction. Adults also considered statements with *or* infelicitous when both disjuncts were true. The study also found that the 2AFC and 3AFC tasks registered different aspects of adult interpretations: the 2AFC task captured adult intuitions on the basic semantics of the connectives while the 3AFC task was sensitive to pragmatic infelicities as well.

Methods.

Materials and Design. The study used six cards with cartoon images of a cat, a dog, and an elephant (Figure 1). There were two types of cards: cards with only one animal and cards with two animals. There were three types of guesses: simple (e.g. *There is a cat*), conjunctive (e.g. *There is a cat and a dog*), and disjunctive (e.g. *There is a cat or a dog*). In each guess, the animal labels used in the guess and the animal images on the card could have no overlap (e.g. Image: dog, Guess: *There is a cat or an elephant*), partial overlap (e.g. Image: Cat, Guess: *There is a cat or an elephant*), or total overlap (e.g. Image: cat and elephant, Guess: *There is a cat or an elephant*). Crossing the number of animals on the card,

the types of guesses, and the overlap between the guess and the card yields 12 different possible trial types. I chose 8 trial types (Figure 2), to balance the number of one-animal vs. two-animal cards, simple vs. connective guesses, and expected true vs. false trials.

Participants and Procedure. I used Amazon's Mechanical Turk (MTurk) for recruitment and the online platform Qualtrics for data collection and survey design. The task took about 5 minutes on average to complete. 109 English speaking adults participated. 57 of them were assigned to a 2AFC judgment task and 52 to a 3AFC judgment task. In the 2AFC task, participants had to judge using the options "wrong" and "right". In the 3AFC task they had to choose between "wrong", "kinda right", and "right". The two conditions were otherwise identical. There are many possible labels for the middle option "kinda right", including "kinda wrong" or "neither". In a later experiment (not reported in this dissertation) I tested different intermediate labels and found that adults consider "kinda right" to be the most suitable option for capturing pragmatic infelicity (Jasbi, Waldon, & Degen, submitted). I expect similar behavior from labels like "a bit right" and "a little right" which refer to non-maximal degrees of being right.

The experiment had three phases: introduction, instruction, and test. In the introduction, participants saw the six cards and read that they would play a guessing game. Then a blindfolded cartoon character named Bob appeared on the screen. Participants were told that in each round of the game, they would see a card and Bob was going to guess what animal was on the card. I emphasized that Bob could not see anything. I asked participants to judge whether Bob's guess was right. In the instruction phase, participants saw an example trial where a card with the image of a dog was shown with the following sentence written above Bob's head: *There is a cat on the card.* All participants, who correctly responded with "wrong", proceeded to the test phase.

In the test phase, participants saw one trial per trial type. Within each trial type, the specific card-guess scenario was chosen at random. The order of trial types was also randomized. At the end of the study, participants received $0.4 as compensation. Figure 3

shows an example test trial.

Table 1

Summary of Study 1 Methods

Study	N	Age	Mode	Response Options
Study 1 - Part 1	57	Adults	Online (Mturk)	Wrong, Right
Study 1 - Part 2	52	Adults	Online (Mturk)	Wrong, Kinda Right, Right

Results. In this section, I first present the results of the 2AFC and 3AFC tasks with adults. Then I discuss how these results can be interpreted with respect to the semantics and pragmatics of disjunction in the context of the guessing game.

Judgments with Two Alternatives (2AFC). Figure 4 shows the results for the adult 2AFC task. The two left columns show the simple guesses and serve as controls. The results show that if the animal mentioned in the guess was not on the card (e.g., elephant), participants judged the guess to be "wrong"; if the animal was on the card (e.g., cat), participants judged the guess to be "right". The next two columns of Figure 4 show the results for the test conditions, namely conjunction and disjunction. They match the expectations for logical conjunction and (inclusive) disjunction: an *and*-guess (e.g. cat and dog) is "wrong" if only one of the animals is shown on the card, and "right" if both are on the card. An *or*-guess (e.g. cat or dog) is "right" whether one or both animals are depicted on the card.

Judgments with Three Alternatives (3AFC). Figure 5 shows the results for the 3AFC judgment task. For four trial types, the results are identical to the 2AFC task: if the animal mentioned in the guess was not on the card (e.g. elephant), participants judged the guess "wrong". If the animal mentioned (e.g. cat) was the only animal on the card, participants judged the guess "right". Finally, if there were two animals and the puppet mentioned them using *and* (e.g. cat and dog), all participants considered the guess "right".

The four remaining trial types showed different patterns of judgments than the ones in

the 2AFC task. If the animal mentioned (e.g. cat) was only one of the animals on the card, participant judgments were divided between "right" and "kinda right" (See Table 2, row 1 for the statistical test). Also, most adults considered a conjunctive guess (e.g. cat and dog) "wrong", when only one of the animals was on the card (Table 2, row 2). However, some considered it "kinda right", perhaps suggesting that the intermediate option was used to express the notion of "partial truth". When both animals were on the card everyone agreed that the conjunctive guess was "right".

With respect to disjunctive guesses like "cat or dog", if the card had only one of the animals, most adults considers the guess "right" while some considered it "kinda right" (Table 2, row 3). It is possible that the adults who considered such guesses "kinda right" were sensitive to the under-informative nature of a disjunctive guess when a simple guess like "cat" would have been more appropriate. If both animals were on the card, adults were split between "kinda right" and "right" responses (Table 2, row 4). The choice of "kinda right" over "right" in such trials can be interpreted as a sign that adults were sensitive to the infelicity of a disjunction when conjunction was more appropriate. However, the scalar reasoning with *and* and *or* is subtle and in section , I discuss the nature of this reasoning in the context of this guessing game.

Table 2

Exact One-Sided Binomial Test

Trial Type	n_{right}/n_{total}	$\hat{p}_{right}$	p_{null}	P-value	95% CI
Two Animals - Simple	32/52	0.62	0.50	0.06	0.49-1
One Animal - AND	16/52	0.69	0.50	0.00	0.57-1
One Animal - OR	19/52	0.63	0.50	0.04	0.51-1
Two Animals - OR	32/52	0.62	0.50	0.06	0.49-1

Discussion. The example sentences bellow show the common interpretations of conjunctive and disjunctive assertions (Aloni, 2016).

- Bob is sad *and* angry.
  - Both are true. (Truth Conditional Meaning)
- Bob is sad *or* angry.
  - At least one of the two is true. (Truth Conditional Meaning)
  - Speaker doesn't know which is true. (Ignorance Inference)
  - At most one of the two is true. (Exclusivity Inference)

A conjunctive assertion implies that both propositions are true while a disjunctive assertion implies that at least one is true. These two inferences follow from the classical truth-conditional account of conjunction and disjunction. They constitute the semantics of *and* and *or*. However, a disjunctive assertion often has two additional inferences: an ignorance inference and an exclusivity inference. These additional inferences are often classified under pragmatic meaning. This section discusses the semantics and pragmatics of *and* and *or* in the context of the guessing game in Study 1.[1]

The Semantics of AND and OR. Let's assume that the semantics of *and* and *or* in simple declarative sentences like "there is a cat or(and) a dog" is captured by the logical operators conjunction and inclusive disjunction respectively. A conjunction is true when both conjuncts are true and false otherwise. An inclusive disjunction is true when at least one disjunct is true and false otherwise. Let's also assume a simple linking function in which false statements are judged as "wrong" and true statements as "right" (see Jasbi et al. (submitted) for a discussion of linking assumptions in this task). In the context of study 1, this purely semantic (i.e. truth-conditional) account has two main predictions: 1. Conjunctive guesses like "cat and dog" are wrong when only one of the animals is on the

[1] See Gutzmann (2014) for a comprehensive discussion of the definitions and boundaries of semantics and pragmatics. Here my definitions and assumptions are close to those of Gazdar (1979).

card. 2. Disjunctive guesses are always right because in all such trials at least one of the animals is present on the card. Figure 4 shows that in 2AFC judgments, both predictions are borne out. In other words, judgments with two alternatives seem to match the predictions of a purely semantic account of the connectives *and* and *or* with a linking function that considers "right" and "wrong" roughly as "true" and "false".

However, in the 3AFC task, judgments deviated from a purely semantic account in four trial types: 1. disjunction trials with one animal 2. disjunction trials with two animals, 3. conjunction trials with one animal, and 4. trials with simple guesses when two animals were shown on the card. Participants often used the third option "kinda right" in these trial types. Other trial types obtained identical results in 2AFC and 3AFC tasks. The comparison of forced choice judgments with two and three alternatives suggests that two alternatives better captured the truth-conditional meaning of the connectives, but underestimated adult pragmatic reasoning in the guessing game.

The Pragmatics of AND and OR. A disjunctive assertion like "cat or dog" gives rise to an ignorance inference and an exclusivity inference. The ignorance inference is the inference that the speaker does not know which disjunct actually holds. For example in figure 6, the disjunctive guess is uncertain between three outcomes: cards 1, 2, and 3. As pointed out in Chapter **??**, a disjunction is infelicitous when the outcome is known to discourse participants. For example, Tarski mentioned that a disjunction like "the grass is green or blue" is odd because we already know that it is green. The guessing game in this study controls for this ignorance effect by keeping the guesser blindfolded. Therefore, all the disjunctive guesses are evaluated in a context where participants know that the guesser is ignorant of the animals on the cards - both the number of them on the card and their identity. The exclusivity inference is the inference that only one of the disjuncts holds and **not both**. In figure 6, a disjunction like "cat or dog" only refers to cards 2 and 3 if it is accompanied by an exclusivity inference.

Since Grice (1989), this exclusive interpretation of *or* has been (at least partly)

attributed to pragmatic reasoning about the speaker's connective choice. The reasoning goes like this: conversational participants are required to make their utterances as informative as possible. In the context of making predictions and guessing, a guesser is required to make any guess as specific (i.e. informative) as possible.[2] A conjunction is more specific and informative than a disjunction (Horn, 1989). For example in Figure 6, "cat and dog" picks card 1 while "cat or dog" refers to cards 1, 2, and 3. If speakers intend to refer to card 1, they should use *and* and say "cat and dog". If they use *or* instead of *and*, they probably do not intend to refer to card 1. Following this line of reasoning, I can exclude the possibility that a speaker intends to refer to card 1. The term "exclusivity implicature" captures this pragmatic reasoning that results in excluding the possibility of both disjuncts being true.

My goal here is to lay out the structure of pragmatic reasoning in the experimental setup and explain how it is manifested in the results of the experimental studies. There are three main components to the pragmatic reasoning in the guessing game: 1. the assumptions of the game. 2. sensitivity to (under)informativity, and 3. the pragmatic reasoning about the speaker's choice of connectives. Like Katsos & Bishop (2011), I have considered "sensitivity to informativeness" as a precondition for "derivation of scalar implicatures". I begin with the assumptions of the guessing game.

- **Guessing Game Assumptions**:
  - **Ignorance**: the guesser does not know the number or identity of the animals on the card.
  - **Specificity**: A guesser is required to be as specific as possible, ideally referring to a single card.

As explained before, ignorance of the guesser was explicit and part of the instructions

[2]When you ask someone to predict the outcome of a coin toss, a guess like "it will be heads or tails" does not count as a felicitous guess or prediction, presumably because it will always be true.

in the study. However, specificity was an implicit assumption[3]. All the guesses used in the experiment can pick a single card except for disjunctive ones. Conjunctive guesses like "cat and dog" pick specific cards. The simple ones like "cat" can be strengthened pragmatically to mean "only a cat", and pick a specific card. However, Disjunctive ones like "cat or dog" pick two cards in their most specific (exclusive) sense. Therefore, they are always under-informative and violate the specificity assumption.

- **Sensitivity to Informativeness**: The guesser said "cat or dog" which is under-informative and picks card 1, 2, and 3.
  - **Violation Assumption**: the guesser is violating the specificity requirement.

Participants can detect the underinformativity of disjunctive guesses, notice the violation of specificity, and then decide whether they would like to tolerate this violation or punish it. It should be pointed out that it is hard to distinguish between "tolerating the specificity violation" and simply revising the specificity assumption of the game to avoid a violation. For example, participants may assume that the goal of the game is saying something true about the cards rather than being as specific as possible. In either case, the prediction is that adults who tolerate violation or revise specificity would judge disjunctive guesses as "right". However, if participants assume specificity and decide to not tolerate its violation, they will judge all disjunctive guesses to have some degree of infelicity. Since an under-informative guess is still technically correct, participants may not punish such a guess with a "wrong" response and prefer an intermediate option like "kinda right". This is what study 1 shows. With two alternatives, not many adults judge infelicity with disjunctive guesses and there are almost no "wrong" responses. With three alternatives, "kinda right" responses pop up. Adult responses are split between "kinda right" and "right".

If detecting and reacting to underinformativity is the whole story, then disjunctive guesses should show similar degrees of infelicity, regardless of how many animals there are on

[3]Making this assumption explicit is both hard for young children and almost impossible when disjunctive guesses are used. Disjunctive guesses are always underinformative and never pick out a specific card.

the card. However, the results of the 3AFC task suggest otherwise. A logistic mixed-effects model with the random intercepts and slopes for subjects and fixed effect of disjunction type found that when comparing disjunctive guesses in the 3AFC task, participants were more likely to choose "kinda right" than "right" when both animals were on the card (β=-1.22, z=-2.25, p=0.02). In other words, participants judged further infelicity with disjunctive guesses that had both disjuncts as true. Therefore, it is possible that when both disjuncts were true, some participants went through the following pragmatic reasoning:

- **Reasoning on Alternatives**: Why did the guesser choose the under-informative connective *or* rather than the more informative *and*?
  - **Resolution Assumption**: speaker is trying to be as specific as possible by resolving the issue of how many animals are on the card.
    - **Exclusivity Implicature**: Given the resolution hypothesis, if the speaker had decided that two animals were on the card, they should have said "cat *and* dog". They did not, so they had decided that only one animal is on the card and **not both**.

How does the exclusivity implicature affect participant judgments in the experimental setting? One possibility is that excluding the correct response pragmatically is treated like cases of excluding the right response semantically. For example, guessing "elephant" when there is a cat on the card. The prediction is that disjunctive trials with true disjuncts should receive "wrong" responses. However, this prediction was not borne out. Such disjunctive trials are almost never judged as "wrong".

Alternatively, it is possible that adults differentiate incorrect pragmatics from incorrect semantics (i.e. falsehood) and punish incorrect pragmatics less than incorrect semantics. This conclusion is supported by the response patterns across trial types (figure 5). Trial types that received a "wrong" response were those that were false. Pragmatically infelicitous trial types, namely simple guesses like "cat" or disjunctive guesses like "cat or dog" when

both animals are on the card, receive "kinda right" responses. In other words, adults consider false utterances as "wrong" guesses but infelicitous utterances do not reach the level of being "wrong"; they are still right even though not completely right. This would explain why the rates of infelicity (avoiding the "right" alternative) differ between 2AFC and 3AFC tasks in disjunctive trials with true disjuncts (0.18 vs. 0.62).

Study 2: Children's three-alternative judgments and open-ended feedback

The goal of this study was to examine children's interpretations of *and* and *or* in the guessing game and compare them to those of the adults. Since the 3AFC judgment task in study 1 proved better at capturing the nuances in adults' pragmatic reasoning, I decided to first test children using three alternatives. I also analyzed children's open-ended comments about the guesses in the experimental context. Both three-alternative judgments and the analysis of children's open-ended responses showed that children differentiate *and* and *or* statements. The judgment task suggested that children do not consider disjunctive guesses with true disjuncts as infelicitous. Yet, the analysis of their open-ended feedback suggests otherwise. Children took issue with such guesses and corrected them. I conclude that the 3AFC judgment task may have underestimated children's pragmatic competence.

Table 3

Summary of Study 2 Methods

Study	N	Age	Mode	Response Option
Study 2	42	3;1-5;2 (M = 4;3)	Study Room	Circle (wrong), Little Star (little right), Big Star (right)

Methods.

Materials and Design. I used the same set of cards and linguistic stimuli as the ones in study 1. There were 8 trial types and 2 trials per trial type for a total of 16 trials. I made two changes to make the experiment more suitable for children. First, instead of the

fictional character Bob, a puppet named Jazzy played the guessing game with them. Jazzy wore a sleeping mask over his eyes during the game (Figure 8). Second, a pilot study showed that a scale with three alternatives is better understood and used by children if it is presented in the form of rewards to the puppet rather than verbal responses such as "wrong", "a little bit right", and "right", or even hand gestures such as thumbs up, middle, and down. Therefore, I placed a set of red circles, small blue stars, and big blue stars in front of the children. These tokens were used to reward the puppet after each guess. During the introduction, the experimenter explained that if the puppet is right, the child should give him a big star, if he is a little bit right, a little star, and if he is not right, a red circle.

Participants and Procedure. I recruited 42 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;1 and 5;2 years old (Mean = 4;3). The experiment was carried out in a quiet room and all sessions were videotaped. There was a small table and two chairs in the room. Children sat on one side of the table and the experimenter and the puppet on the other side facing the child. The groups of circles, small stars, and big stars were placed in front of the child from left to right. A deck of six cards was in front of the experimenter. As in study 1 with adults, the children went through three phases: introduction, instruction, and test.

The goal of the introduction was for the experimenter to show the cards to the children and make sure they recognized the animals and knew their names. The experimenter showed the cards to the children and asked them to label each animal. All children recognized the animals and could label them correctly. In the instruction phase, children went through three example trials. The experimenter explained that he was going to play with Jazzy (the puppet) first, so that the child could learn the game. He removed the six introduction cards and placed a deck of three cards face-down on the table. From top to bottom (first to last), the cards had the following images: cat, elephant, cat and dog. He put the sleeping mask on the Jazzy's eyes and explained that the puppet is going to guess what animal is on the cards. He then picked the first card and asked the puppet: "*What do you think is on this card?*"

Jazzy replied with "*There is a dog*". The experimenter showed the cat-card to the child and explained that when the puppet is "not right" he gets a circle. The pilot study had shown that some children struggle with understanding the word "wrong", so "not right" was used instead. He then asked the child to give the puppet a circle. Rewards were collected by the experimenter and placed under the table to not distract the child. The second trial followed the same pattern except that the puppet guessed "right" and the experimenter invited the child to give the puppet a big star. In the final trial, the puppet guessed that there is a cat on the card when the card had a cat and a dog on it. The experimenter said that the puppet was "a little right" and asked the child to give him a little star.

Table 4

Instruction Trials.

Card	Guess	Reward
CAT	There is a cat!	Circle
ELEPHANT	There is an elephant!	Big Star
CAT-DOG	There is a dog!	Little Star

In the test phase, the experimenter removed the three instruction cards and placed a deck of 16 randomized cards on the table. The experimenter explained that it was the child's turn to play with the puppet. The test phase followed the pattern described in the instruction phase. The randomization code as well as the details of the methods are in the online repository for this study at https://github.com/jasbi/jasbi_dissertation_LearningDisjunction.

Offline Annotations. During analysis of the videos, children's linguistic feedback to the puppet after each guess was categorized into four types: 1. None, 2. Judgments, 3. Descriptions, and 4. Corrections. The first category referred to cases where children did not say anything and only rewarded the puppet. Judgments referred to linguistic feedback such as *you are right!*, *yes*, *nope*, or *you winned*. Such feedback only expressed judgments and

complemented the rewards. Descriptions were cases that the child simply mentioned what was on the card: *cat!*, *dog and elephant!*, *There is a cat and a dog!* etc. Finally, corrections referred to feedback that provided additional linguistic elements that acted like corrections to what the puppet had said. Examples include: *Just a cat!*, *Both!*, *The two are!*, *Only cat*, *cat AND dog* (with emphasis placed on *and*). In trials where the child provided both judgments as well as descriptions or corrections, I placed the feedback into the more informative categories, namely description or correction.

Results. Figure 9 shows the results for children's 3AFC judgments. Starting from the left column, if the mentioned animal was not on the card (e.g. elephant), children judged the guess as "wrong". If the animal mentioned (e.g. cat) was the only animal on the card, children judged the guess to be "right". Here I will ignore the results for trial types in which the animal mentioned was one of the animals on the card. The reason is that such trials were used in the instruction phase to introduce the "little bit right" guesses, and the results are potentially biased by the instructions.

In conjunctive guesses (e.g. *cat and dog*), when only one of the animals mentioned was on the card, children judged the guess as "wrong" or "a little bit right". However, if both animals were on the card, they judged the conjunctive guess as "right". In disjunctive guesses (e.g. *cat or dog*), when only one of the animals mentioned was on the card, children considered the guess "right" or "kinda right". If both animals were on the card, the disjunctive guess was considered "right".

The comparison of conjunction and disjunction trials (last two columns of figure 9) shows that overall, children distinguished between *and* and *or* when one animal was on the card. Given that the one-animal conjunction trials are false but the one-animal disjunction trials are true, the difference in response patterns may suggest that children understood the truth-conditional differences between *and* and *or*. The truth judgments did not provide evidence that children differentiated *and* and *or* when two animals were on the card. Since in the majority of examples with *or* and two animals, children responded with "right", it is

possible to conclude from the judgment data that children did not generate exclusivity inferences in this task.

Figure 10 compares the results for children and adults' 3AFC judgments in the conjunction and disjunction trials. The major difference between adults and children's responses was disjunctive trials with two animals on the card. Most children considered such trials as "right" while adults considered them as "kinda right" or "right". In the next section, I use Bayesian regression modeling to compare adults' and children's three-alternative responses more systematically.

Analysis and Statistical Modeling. I used the R package RStan for Bayesian statistical modeling to fit separate ordinal mixed-effects logistic models for the children's and adults' judgments. The response variable had three ordered levels: *wrong*, *kinda right*, and *right*. The trial types *One-Animal-OR*, *Two-Animals-OR*, *One-Animal-AND* constituted the (dummy-coded) fixed effects of the model with *Two-Animals-AND* set as the intercept. The model also included by-subject random intercepts. The priors over trial types and the random intercepts were set to $\mathcal{N}(0, 10)$. I also included parameters C_1 and C_2, the two cutpoints delimiting the logistic for 1) *wrong* and *kinda right* and 2) *kinda right* and *right* responses, drawn with the prior $\mathcal{N}(0, 1)$.[4] All four chains converged after 3000 samples (with a burn-in period of 1500 samples).

I made inferences based on the highest-posterior density (HPD) intervals for the coefficients estimated from each model. Because predictors are dummy-coded, it's possible to examine contrasts of interest by computing the difference between coefficients for pairs of conditions I wish to contrast. In naming the coefficients like b(OR,2), OR/AND represents the connective used and the number represents the number of animals on the card. Figure 11 shows the contrasts of interest: *b(OR, 2)-b(OR,1)* represents the difference between the estimated coefficients for the disjunction trials with two animal on the card and those with only one; *b(OR, 2)* represents the difference between the estimated coefficients for the

[4] I used a tight prior in this case to decrease posterior correlations between cutpoints and intercept.

conjunction trials with two animals and the disjunction trials with two animals; and so on.

Overall, adults' and children's estimated coefficients are similar in sign to one another, though adults' are more extreme. In the conjunction trials (*b(AND, 2)-b(AND,1)*), children and adults showed a strong preference for the cards with two animals rather than one. At the same time, given two animals on the card, children and adults showed a preference for *and* rather than *or* (*b(OR, 2)*). However, with only one animal on the card, children and adults preferred a disjunctive guess (*b(OR, 1)-b(AND,1)*). These results are compatible with the truth conditions of conjunction and disjunction.

The main difference between adults and children shows up in the contrast between the disjunctive trial types: two animals vs. only one (*b(OR, 2)-b(OR,1)*). On average, children rated disjunction trials with two animals higher than those with only one. Adults on the other hand showed the opposite pattern: they rated disjunction trials with two animals lower. This pattern is compatible with current accounts of pragmatic development that suggest an absence of implicatures in children's interpretations. The idea is that while adults strengthen the disjunctive guess "cat or dog" to "cat or dog but not both", children simply interpret it as "cat or dog or both". Adults are therefore going to rate trials with both disjuncts true lower.

The slight preference children show for cards with two animals when the guess is disjunctive is also compatible with the account proposed by Singh et al. (2016) and Tieu et al. (2016). However, the effect is much smaller here than was reported in their studies. The comparison with conjunction trials makes it clear that overall, children are not interpreting *or* as having a conjunctive meaning. The effect in this study can be more accurately described as a preference in judgment for both disjuncts being true rather than a conjunctive interpretation of disjunction. The results from children's spontaneous linguistic feedback make it less likely that children interpretive *or* as a conjunction. I will discuss this issue further in section .

Table 5

Definitions and Examples for the Feedback Categories.

Category	Definition	Examples
None	no feedback provided to the puppet, only reward	
Judgment	the child said yes/no, you are right, etc.	"No!" , "You are right Jazzy!"
Description	mentioned the animal(s) on the card	"elephant", "cat and dog"
Correction	used focus particles like *only/just*, emphasized *and* or used *both*	"only cat", "just elephant", "both!", "cat AND dog!"

Children's open-ended feedback. As explained in section , I also categorized and annotated children's spontaneous and free form verbal reactions to the puppet's guesses. Table 5 summarizes the definitions and examples for each category and Figure 12 shows the results.

I should first point out that each trial type has a similar number of "None" cases, for feedback. Some children remained more or less silent throughout the experiment and only provided rewards to the puppet. In the next study I focus on children's open-ended feedback. In the discussion and analysis here I will not comment further on the "None" category but focus on the other three categories.

In the leftmost column, when the animal guessed (e.g. elephant) was not on the card, children either provided judgments like "No!" or described what was on the card like "cat" or "cat and dog". However, when the animal guessed was the only animal on the card (e.g. cat), most children provided a positive judgment like "Yes". When the animal guessed was only one of the animals on the card, children described what was on the card, say, "cat and dog". Corrections were rare for all these four control trial types.

In the critical trial types with conjunction and disjunction, children showed a high rate

of corrections and description when the guess used *and* but there was only one animal on the card. In their corrections, children used the focus particles *just* and *only* like "just a cat" or "only a cat". However, in trial types where conjunction was used and both animals were depicted, children predominantly provided positive judgments like "Yes!" and "You are right". Considering disjunctive guesses like "cat or dog", when only one of the animals was on the card, most children simply described what was on the card, for example "cat". However, when both animals were on the card, children corrected the puppet by saying *both* or emphasizing *and* as in "cat AND dog".

I performed chi-squared goodness-of-fit tests to compare the feedback distributions in the critical conditions with *and* and *or*. Here I focus on those trials (the four bar charts on the right of Figure 12). Children's linguistic feedback showed three patterns. First, the one-animal conjunctive and two-animal disjunctive (top left and bottom right) trials contained a higher proportion of Corrections than the other trial types. These were trials where the guesses were either false or infelicitous. In the conjunction trials, a comparison of the feedback distribution in one-animal and two-animal conditions was statistically significant ($\chi^2(3, 83) = 201.65$, $p < .0001$), suggesting that children gave different feedback to true and false guesses. A similar numerical trend was present in the disjunction trials, but it was not significant ($\chi^2(9, 4) = 12$, $p = 0.21$).

Second, the one-animal disjunctive trials (top right) showed the highest proportion of *Descriptions*. These are trials in which the guess is correct but not specific enough: it leaves two possibilities open. These trials were significantly different from the one-animal trials for conjunction ($\chi^2(3, 83) = 62.16$, $p < .0001$). Finally, the two-animal conjunctive trials (bottom left) showed the highest proportion of *Judgments* such as *You are right!*. This was not surprising given that these trials represented the optimal guessing scenario. These trials had a significantly different feedback distribution from the matching disjunction trials ($\chi^2(3, 84) = 184.98$, $p < .0001$).

Discussion. In study 2, I used a 3AFC judgment task to test children's comprehension of logical connectives *and* and *or*. I compared these results to those found in the 3AFC judgment task of study 1 with adults. The general comparison showed that adults and children had similar patterns of judgments, except when both disjuncts were true. In such cases, adults judged the disjunctive guess as not completely right while most children found it completely right. There was even a slight preference among children to rewarded the puppet more in such cases, compared to cases of disjunction when only one disjunct was true.

To consider another measure of children's comprehension, I also looked at children's spontaneous open-ended feedback in response to the guesses. My analysis suggested that children recognized false and infelicitous utterances with the connectives and provided appropriate corrective feedback. As expected from an adult-like understanding of connectives, children corrected the puppet most often when there was only one animal on the card and the guess was conjunctive, or when there were two animals on the card and the guess was disjunctive. Perhaps the most important finding was that children increased their corrective feedback in disjunctive guesses where both disjuncts were true, compared to those with only one true disjunct. These findings differ from the results of the 3AFC judgment task which suggested that children did not find any infelicity with disjunctive guesses when both disjuncts were true.

The analysis of children's open-ended feedback raises two important issues. First, as I mentioned before, it runs counter to what the 3AFC judgment task suggests with respect to exclusivity implicatures (i.e. trials with disjunction when both disjuncts are true). The forced-choice judgment task suggests that children find such cases unproblematic while analysis of their spontaneous feedback shows that they provided more corrections to the puppet. Second, a common explanation for why children fail to derive0 implicatures is that they cannot access the stronger alternative to the disjunction *or*, namely *and* (Barner, Brooks, & Bale, 2011). However, in the context of the guessing game, some children explicitly mentioned the word *and* as what the puppet should have said instead of *or*.

Interestingly, these children continued to reward the puppet and considered the guess "right". This raises the possibility that children's forced-choice truth value judgments, whether with two or three alternatives, do not fully reflect their pragmatic knowledge. In study 3, I used both a 2AFC truth judgment task and an analysis of children's open-ended feedback. If the findings of study 2 were on the right track, I expected to replicate the same pattern in study 3: the analysis of children's spontaneous linguistic feedback provides more evidence that children are sensitive to pragmatic violations than the results of the 2AFC judgments.

Study 3: Children's 2AFC judgments and open-ended feedback

This study used the same paradigm as study 2 but focused on children's open-ended feedback and aimed at replicating the findings in study 2. The main hypothesis was that four-year-olds provide corrective feedback to the puppet if both disjuncts are true, but they do not consider this infelicity to be grave enough to render the guess itself "wrong". The main hypothesis along with relevant analyses and predictions were preregistered in an "As Predicted" format[5]. The study used a 2AFC judgment task to compare with the open-ended feedback results. The prediction was that children would provide corrective feedback to the puppet when both disjuncts were true, yet consider the guess "right" and not reflect this infelicity in their truth value judgments. This is what the study found.

Table 6

Summary of Study 1, 2, and 3 Methods

Study	N	Age	Mode	Response Options
Study 1 - Part 1	57	Adults	Online (Mturk)	Wrong, Right
Study 1 - Part 2	52	Adults	Online (Mturk)	Wrong, Kinda Right, Right

[5]The As Predicted pdf document is accessible at https://aspredicted.org/x9ez2.pdf.

Study	N	Age	Mode	Response Options
Study 2	42	3;1-5;2 (M = 4;3)	Study Room	Circle (Wrong), Little Star (Little Right), Big Star (Right)
Study 3	50	3;6-5;9 (M = 4;7)	Study Room	Yes (Right)/No (Wrong) - Open-ended Feedback

Methods.

Materials and Design. Study 3 was similar to Study 2 but differed in how children provided their judgments. Based on the findings in Study 2, I focused on verbal judgments and feedback, instead of rewards. I used two different ways of measuring children's judgments. First, I encouraged children to provide verbal feedback to the puppet. They were asked to say "yes" when the puppet was right, and "no" when he was not. They were also encouraged to help the puppet say it better when he was not right. After children were done with this initial open-ended feedback, for each trial I asked a forced choice yes/no judgment question: "Was Jazzy (the puppet) right?". This question elicited a "yes" or "no" response for each trial independent of their earlier open-ended response. These two measures allowed me to compare open-ended and forced-choice judgments.

Participants and Procedure. I recruited 50 English speaking children from the Bing Nursery School at Stanford University. Children were between 3;6 and 5;9 years old (Mean = 4;7). The setup and procedure were similar to Study 2, except there were no rewards on the table. As before, participants sat through three phases: introduction, instruction, and test. The introduction phase made sure children knew the names of the animals on the cards. In the instruction phase, they received four training trials, as shown in Table 7.

As in Study 2, the experimenter put a sleeping mask over the puppet's eyes and explained that Jazzy (the puppet) was going to guess what animal was on the cards. He

then picked the first card and asked the puppet: "*What do you think is on this card?*" The puppet replied with "*There is a dog*". The experimenter showed the cat-card to the child and said: when Jazzy is *not right*, tell him "no". He then asked the child to say "no" to the puppet. The second trial followed the same pattern except that the puppet guessed *right* and the experimenter invited the child to say "yes" to the puppet. There were two more instruction trials before the test phase began. This contained 16 randomized trials, half of which contained guesses with the words *and* and *or*. The randomization code as well as the details of the methods are on the online repository for this dissertation at https://github.com/jasbi/jasbi_dissertation_LearningDisjunction.

Table 7

Instruction Trials for Study 3.

Card	Guess	Response
CAT	there is a dog!	No!
ELEPHANT	there is an elephant!	Yes!
DOG-ELEPHANT	there is a cat!	No!
DOG	there is a dog!	Yes!

Results. I first look at the results of the 2AFC judgement task for each trial type and compare them to those of the adults' in Study 1. Then I analyze children's open-ended responses and compare them to the forced choice responses obtained in the same trial types. For the 2AFC judgments I excluded 26 trials (out of total 800) where children either did not provide a Yes/No response or provided both (i.e. "Yes and No"). The exclusions were almost equally distributed among different types of guesses and cards. In the analysis of children's open-ended feedback, I excluded 8 trials (out of total 800) where children either did not provide any feedback or their feedback could not be categorized into the existing categories.

Two-Alternative Forced Choice Judgments. Figure 13 shows children's 2AFC judgments. In the leftmost column, when the animal guessed was not on the card

(e.g. elephant), children considered the guess "wrong". When the animal guessed was the only animal on the card (e.g. cat), children considered the guess "right". However, if the animal guessed (e.g. cat) was only one of the animals on the card, children were equally split between "wrong" and "right" judgments. On the other hand, almost all adults considered such guesses "right" in their 2AFC judgments (Figure 4). In such trial types, children seem to interpret the guess "there is a cat" as "there is **only** a cat", while adults do not. This difference between children and adults is unexpected for a theory of meaning acquisition that assumes children are overall more logical or literal as interpreters than adults (Noveck, 2001).

In the trials with *and* and *or*, children's judgments were similar to those of adults. Figure 14 compares adults' and children's 2AFC judgments. In trials with conjunction, when only one of the animals was on the card, most children considered the guess "wrong". This is similar to adults' judgments, but different in extent: adults were more consistent and unanimous in rejecting such guesses. A mixed effects logistic regression with the fixed effect of age category (adult vs. child) and random effect of subject found no significant difference between adults' and children's responses in such trials (see Table 8, Conjunction - One Animal).

Table 8

Mixed effects logistic models for conjunction and disjunction trials when only one disjunct was true, in 2AFC judgments of adults and children, using `glmer` *in R's lme4 package. Formula:* $Response \sim AgeCategory + (1|Subject)$.

Trial Data	Coefficient	Standard Error	Z-Value	P-value
Conjunction - One Animal	-2.05	2.86	-0.72	0.47
Disjunction - One Animal	1.34	1.79	0.75	0.45

In conjunctive guesses where both animals were on the card, both children and adults were unanimous in considering the guess "right". In disjunctive trials when only one of the animals was on the card, most children considered the guess "right". This is again similar to

adults but differs from them in extent: adults more consistently and unanimously judged such guesses as "right". Yet again, a mixed effects logistic regression with the fixed effect of age (adult vs. child) and random effect of subject found no significant difference between adults' and children's responses in such trials (see Table 8, Disjunction - One Animal). Adults and children showed almost identical patterns of judgments in trials where there was two animals on the card and the guess used the connective *or*. Children and adults did not differ in their rate of rejecting disjunctive guesses when both disjuncts were true.

Finally, there is a small but significant preference in children's judgments of disjunctive statements for both disjuncts to be true. Comparing the disjunctive trials with one animal and two animals on the card, a mixed-effects logistic model with the fixed effect of disjunction type and the random effect of subjects found that children had a slight preference for both animals to be on the card (b= 1.85, se= 0.56, z= 3.32, $p < 0.001$). There was a similar small trend in children's three-alternative judgments in study 2. While this was quite small compared to the other effects observed in these studies, it nevertheless indicated a difference between children's and adults' judgments. I return to this in more detail in section of the General Discussion.

Open-ended Feedback. Figure 15 shows the distribution of children's feedback to the puppet in Study 3 (see Table 5 for the definitions and examples of feedback categories). There were no "None" responses in this study since the experimenter explicitly asked children to provide feedback to the puppet. The distribution of the responses in the other three categories (Judgment, Description, and Correction) revealed a successful replication of Study 2.

Children's feedback showed four main patterns. First when the puppet guessed an animal not on the card (e.g. elephant), there is a split pattern between negative judgments like "no!" and simple descriptions of what was on the card, e.g. "cat!". Children provided no corrections on such trials. Second, almost all children responded with positive judgments like "yes" when the puppet's guess accurately matched what was on the card. This was the case

in trials where there was only one animal on the card (e.g. cat) and the puppet mentioned it, as well as trials where there were two animals on the card and the puppet mentioned both with a conjunction (e.g. cat and dog). Third, children provided the largest number of corrective feedback in trials where the guess was either false or infelicitous. These included three trial types: (a) the ones where there were two animals on the card but the puppet only guessed one (e.g. cat); (b) the ones where the puppet guessed two animals with conjunction (e.g. cat and dog) but only one of them was on the card; and (c) the ones where there were two animals on the card, and the puppet guessed both but used a disjunction (e.g. cat or dog). The fourth general pattern was unique to disjunctive trials with only one animal on the card. In such cases, almost all children simply named the animal on the card (e.g. "Cat!").

Figure 16 breaks down children's open-ended feedback based on whether children said "yes", "no", or said something else. Responses that were not "yes/no" judgments are grouped in a middle category shown with a dash. The goal here is to compare children's open-ended judgments with their forced choice judgments shown in Figure 13. Children's open-ended judgments and their forced choice judgments in study 3 show similar patterns for all types of guesses except for disjunctive ones. In trials that the puppet guessed with *or*, the vast majority of children refused to provide a "yes/no" judgment when they were not forced to. Instead, they described the animal on the card or provided corrections to the puppet's infelicitous disjunctive guess.

One way to interpret these results is that disjunctive guesses (with at least one disjunct true) are considered neither right nor wrong by almost all children. When children were forced to provide wrong-right responses in the experimental context, some conformed to the adult patterns of judgment and some did not. However, it is possible that such deviations from adult judgments do not reflect differences in the comprehension of disjunction, but rather differences in how children map their adult-like comprehension onto the notions of "right" and "wrong" in a forced choice judgment task. In other words, it is possible that children and adults only differ in how they behave when they are forced to respond with a

fixed set of options.

Figure 17 is similar to Figure 16 but it uses color-fill to show what children said in addition to "yes/no". The gray color represents the trials where children only said "yes/no" and nothing else. The yellow-fill represents descriptions where children mentioned the animal on the card (e.g. "cat!"). The blue fill represents children's corrective feedback that used the exclusive focus particles *just* or *only* (e.g. "just a cat!"). Such a corrective feedback suggests that the guess included an animal that did not belong and should have been excluded. Finally the red fill represents the inclusive corrective feedback that emphasized the word *and* or said *both* (e.g. "Both!", "cat AND dog!"). Such corrective feedback indicated that both animals should have been mentioned.

As shown in the leftmost column, when the puppet mentioned an animal that was not on the card (e.g. elephant), children responded with a simple "no" or "no" followed by what was on the card (e.g. no! elephant!). When there was only one animal on the card and the puppet mentioned the animal (e.g. cat), children responded with a simple "yes". However, when the card had two animals (e.g. cat and dog) and the puppet only mentioned one of them (e.g. cat), children were likely to provide inclusive feedback. They said *both* or emphasized *and*, as in "cat AND dog". However, in such trials children were equally split between saying "yes", "no", or neither.

In the trials with conjunctive and disjunctive guesses, when there was only one animal on the card (e.g. cat) and the puppet used a conjunction (e.g. cat and dog), children were likely to say a simple "no" or say "no", followed by "only/just" (e.g. no, just a cat). Some children did not say "no" but did respond with "only/just". When the card had two animals on it and the puppet mentioned both using *and*, children responded with a simple "yes". In trials with disjunctive guesses like "cat or dog", children avoided yes/no responses. Instead, when the card had only one animal, children mentioned that animal. When the card had both animals, children said *both* or emphasized the word *and*, as in "cat AND dog".

Figure 18 shows the same feedback data in Figure 17 but uses the x-axis to also show

the proportion of feedback categories other than yes/no judgments. My goal here is to display the trial types with corrective feedback (blue and red). These trial types include: (1) conjunction when only one conjunct is true, (2) disjunction when both disjuncts are true, and (3) simple guesses (e.g. "there is a cat") when two animals were on the card. These trial types involved guesses that are either false or infelicitous. Furthermore, the type of corrective feedback children provided matched the type of mistakes made in the guesses. With conjunctive guesses like "cat and dog" when there was only one animal on the card (e.g. cat), children provided exclusive corrections (e.g. just/only a cat), suggesting that the other animal (e.g. dog) should have been excluded. When two animals were on the card and the puppet used a disjunctive guess like "cat or dog", or simple guess like "cat", children provided inclusive feedback, suggesting that another animal should have been included. This is particularly notable in the case of disjunction since both animals were mentioned, but children still emphasized that the connective *and* should have been used, or that both animals mentioned were on the card.

Discussion. Study 3 measured children's comprehension of logical connectives by asking them to judge a puppet's guess in two ways: with open-ended feedback and with a two-alternative forced choice task. First, I asked children to say "yes" to the puppet if he was right and "no" if he was wrong. However, children could provide any form of feedback they wanted. Second, I followed children's open-ended feedback with a two-alternative forced choice question: "Was the puppet right?" This way, I could measure children's comprehension in two different ways in the same trial. Ideally, both measures should show similar results. However, the findings were similar for conjunctive guesses, but not disjunctive ones. Children avoided binary right/wrong feedback with disjunction and preferred to provide more nuanced feedback when they could.

The 2AFC responses followed the predicted pattern: conjunctive guesses were judged wrong if only one conjunct was true, and right if both were true. Disjunctive guesses were judged right whether one or both disjuncts were true. There was no significant difference in

the 2AFC task between the responses of children and those of adults in Study 1.

Children's open-ended feedback in Study 3 replicated the findings of Study 2. Children provided more corrective feedback in false and infelicitous trials than in true and felicitous ones. The corrective feedback was tailored to the puppet's mistake. If the puppet used a conjunction when there was only one animal on the card, children pointed out that the other animal should have been excluded using the exclusive adverbials *just* and *only*. If the puppet used a disjunction when both animals were on the card, children stressed *and* or *both*, implying that both animals should be included.

While the 2AFC results suggested that children take no issue with disjunctive guesses when both disjuncts are true, the analysis of their corrective feedback showed that they provide appropriate corrections in such cases and emphasize that the connective *and* would have been a better guess. Taking both measures together, I conclude that even though children are aware of the problem with such guesses, they do not consider them *wrong*. These results are similar to those I reported for adults in Study 1.

General Discussion

I reported three studies on adults and four-year-olds' comprehension of the logical connectives *and* and *or*. The first study used two- and three-alternative forced choice judgment tasks with adults. In the 2AFC task, adult interpretations closely matched the semantic accounts of *and* and *or* as conjunction and inclusive disjunction. The judgments did not register robust signs of pragmatic infelicities. However, the 3AFC judgment task, showed signs of pragmatic infelicities, especially in disjunctive guesses with true disjuncts. When both disjuncts were true, participants were more likely to choose "kinda right" rather than "right".

The second study used a 3AFC judgment task with four-year-old children. It also included an exploratory analysis of children's open-ended verbal feedback to the puppet in the experimental setting. Children's interpretations were similar to those of adults in the

3AFC task and only differed for pragmatically infelicitous disjunctions. When both disjuncts were true, adults tended to judge disjunctive guesses as "kinda right". This was evidence for the pragmatic infelicity of such guesses. While, children judged such disjunctive statement as "right", the analysis of their open-ended feedback showed that they took issue with such statements as well, and provided appropriate corrective feedback.

In the third study, I focused on eliciting open-ended verbal feedback from children and followed it with a 2AFC task. In the 2AFC task, children's responses reflected the semantics of connectives as conjunction and inclusive disjunction. There was no significant difference between children and adults in the two-alternative judgments. Since the 2AFC task appeared to be a good indicator of semantic knowledge, it seemed reasonable to conclude that adults and four-year-olds displayed similar semantic knowledge of the connectives. Analysis of the children's open-ended feedback replicated the findings in study 2. Children provided more corrective feedback in false and pragmatically infelicitous trials with logical connectives than in felicitous trials. The comparison of the 2AFC task and children's open-ended responses shows that children are sensitive to the infelicity of disjunctions with true disjuncts, even though they consider them to be "right" guesses.

Overall, I did not find any major differences between adults' and four-year-old children's interpretations of logical connectives *and* and *or* in the context of the guessing game. However, there were two minor differences. First, I found that in both 2AFC and 3AFC judgment tasks, children showed a small preference for disjunctions with both disjuncts true rather than only one. Adults on the other hand showed the opposite pattern: they preferred disjuncts with only one disjunct true. Second, in both 2AFC and 3AFC judgment tasks, children rated disjunctions with both disjuncts true higher than adults did. That is, they considered utterances like "there is a cat or a dog" when both animals were on the card "right" more often than adults. Here I will discuss these two differences and their possible causes in more detail.

Preference for True Disjuncts. First for some children, there was a small preference for both disjuncts being true, compared to only one. This effect is similar in kind but not magnitude, to an effect that Singh et al. (2016) and Tieu et al. (2016) reported. In my study this effect is quite small while Singh et al. (2016) and Tieu et al. (2016) seem to have found bigger effects. Based on this, Singh et al. (2016) proposed that many children at this age-range have a pragmatically driven conjunctive interpretation of disjunction. In short, due to a non-adult like alternative set to the connective *or*, children strengthen a disjunctive statement pragmatically and derive a conjunction. The studies reported here provide no support for this proposal. In both 2AFC and 3AFC judgments, children clearly differentiated between disjunctive and conjunctive guesses. Furthermore, analysis of children's open-ended feedback showed distinctly different response patterns for conjunction and disjunction. More importantly, the open-ended feedback to disjunctive guesses showed the opposite pattern to that predicted by the conjunctive hypothesis. Children took issue with disjunctions that had both disjuncts true and provided more corrective feedback in such cases. Therefore, the findings from Singh et al. (2016) and Tieu et al. (2016) appear to be a product of experimental design rather than a real reflection of children's comprehension of the connectives.

However, even if this small preference for true disjuncts is not due to the method of measurement, it can be accounted for in several other ways that have not yet been successfully ruled out. First, the conjunctive interpretation may not be due to a faulty pragmatic computation, but rather a default conjunctive interpretation when the connective is not properly heard, understood or is unknown. To check this hypothesis, it should be possible to test children's comprehension of novel or noisy connectives. A novel coordination like *cat dax dog* with *dax* as a nonce connective could well be interpreted as a conjunction. Such a result would suggest that in studies with high cognitive demand, children may default to a conjunctive interpretation if they miss the relevant connective. Second, the conjunctive preference could be due to some children's preference for the linguistic labels to match the

animals on the card (or more generally a match between linguistic description and the state of the world). This hypothesis is consistent with the results in the other trial type that had a mismatch in the number of animals and the guess, where the guess was still technically true: simple guesses (e.g. there is a cat) with two animals (e.g. cat and dog). Children were equally split between "wrong" and "right" in their judgments here, while adults considered such guesses "right". In light of these alternative explanations, I am hesitant to attribute this small preference to a pragmatically driven conjunctive interpretation of disjunction.

Lack of Infelicity with True Disjuncts. The second difference between adults and children emerged in the 3AFC judgment task: in disjunctive trials (e.g. "cat or dog") with two animals (e.g. cat and dog), adults were more likely to choose "kinda right" than children were. Children mostly chose "right". This response pattern has been taken to mean that children found no infelicity with such disjunctions or that they did not "derive an exclusivity implicature". The absence of an infelicity/implicature is consistent with the generalization that children are more likely than adults to interpret scalar terms literally, and that children do not compute implicatures or judge infelicity to the same **rate** that adults do (Pouscoulous & Noveck, 2009, Katsos (2014)). But why is that?

There have been three major proposals to account for children's low rate of implicatures: 1. processing (Pouscoulous, Noveck, Politzer, & Bastide, 2007; Reinhart, 2004) 2. non-adult-like lexical entry (Barner et al., 2011; Horowitz, Schneider, & Frank, 2017) and 3. pragmatic tolerance (Katsos & Bishop, 2011). Here I show that none of these accounts can provide a satisfactory explanation of the results in this study.

1. Processing. First, the processing accounts locate the problem in children's processing capacities such as working memory. They suggest that pragmatic computations are cognitively taxing and children lack the appropriate processing resources to carry them out appropriately. A prediction of processing accounts (at least in their current format) is that children will show reduced implicature computations for all types of implicatures – scalar or ad-hoc. This prediction was not borne out in my experimental results. In Study 3,

children were much more likely than adults to call a simple guess (e.g. "cat") "wrong" in the 2AFC task if there were two animals on the card (e.g. "cat and dog"). Processing accounts do not predict that children would derive more implicatures than adults but this is what I found for the traditional interpretation of the judgment task.

2. Non-adult-like Lexicon. Several proposals blame the structure of the child's lexicon for the alleged failure in deriving implicatures. The assumption is that the child's lexical entry for scalar items must include three elements for successful derivation: 1. the semantics of the weak term (e.g. *some*, *or*) 2. the semantics of the strong term (e.g. *all*,*and*); and possibly 3. a scale that recognizes the stronger term as an alternative to the weaker one (e.g. <*some*, *all*>, <*or*, *and*>). Each of these elements have been pinpointed as the source of the problem in previous studies (Barner et al., 2011; Horowitz et al., 2017; Katsos & Bishop, 2011). However none of them seem to apply to the results reported here.

If children in this study lack the semantics of the connective *or*, I would expect them to either perform at chance or default to a conjunctive interpretation. Neither prediction was borne out in studies 2 and 3. Furthermore, children's free-form linguistic feedback suggests good understanding of disjunction. So this explanation seems unlikely. The problem cannot be that children do not know the meaning of *and* either. Children's performance in both study 2 and 3 for conjunction trials show that they understand its meaning very well. Finally, while it is possible that children lacked the appropriate lexical scale and could not access the stronger alternative, this explanation cannot be the whole story. Several children in both studies stressed the word *and* suggesting that the puppet should have used the stronger term instead. However, they judged the puppet's guess as "right". If children could not access the stronger term, they would not have mentioned it in their feedback either.

3. Pragmatic Tolerance. Katsos & Bishop (2011) suggested that children tend to tolerate pragmatic infelicities more than adults. They showed that when children are provided with a 2AFC judgment task, they tolerate the infelicity of *some* when *all* applies but when they are presented with a 3AFC task they choose the middle option and report this

infelicity. As in a processing account, the pragmatic tolerance account predicts that scalar and ad-hoc implicatures will be similarly affected. However, my results did not match those of Katsos & Bishop (2011). When children were presented with a 3AFC task, they chose the highest reward (and not the middle option) for uses of *or* when *and* applied. Second, and more importantly, I found different patterns for ad-hoc and scalar implicatures as mentioned before. This is not predicted by the tolerance account unless children are assumed to be more tolerant towards violations of scalar implicatures than they are towards ad-hoc ones. While tolerance may not be the source of the problem here, I believe that a number of discussions including those by Katsos & Bishop (2011) and later Katsos (2014) do point to an important factor: the role of measurement in estimates of children's pragmatic capacity.

Several observations in the current studies provide support for the hypothesis that methodological issues, and more specifically issues of measurement contribute to the differences found between adults and children in pragmatic capacity. First, Study 1 showed that even for adults, the estimates of adult infelicity rates may differ based on the number of alternatives in the forced choice task. A 2AFC task tends to underestimate adults' rate of response to pragmatic infelicity. Second, children's open-ended linguistic feedback in the experimental context better reflected their sensitivity to pragmatic nuances than the forced-choice judgment tasks. Third, children showed a higher rate of infelicity judgments for cases of ad-hoc implicatures (simple guesses with two animals on the card) than adults did. While a difference in sensitivity to ad-hoc vs. scalar implicatures has been reported and argued for before (Horowitz et al., 2017; Stiller, Goodman, & Frank, 2015), a higher sensitivity than adults is not predicted by any of the current accounts.

In order to better understand the differences between adults and children's pragmatic capacities, it is necessary to have a good understanding of how measurements affect estimates of adults and children's performance in the experimental tasks. Children may be no more capable of making exhaustive inferences than adults and no less capable of making scalar inferences either. They may simply have a different construal of the wrong-right scale

and of what the forced-choice task is about. The concepts "right" and "wrong" are as much subject to developmental change and differences between adults and children as are scalar items in general. It is reasonable to assume that children's understanding of what constitutes as "right" or "wrong" does not fully conform to that of adults. However, it remains to be established what these differences are and how they affect the estimates of children's pragmatic abilities. It is important to point out that such issues of measurement could be the culprit behind both children's seemingly slight preference for true disjuncts described earlier and the lack of infelicity judgments when both disjuncts are true.

A General Approach for Measuring Implicature/Infelicity Rate

Methodological issues are nothing new in studies of children's semantic and pragmatic development. Developing better measures of children's linguistic capacities has always been a major concern for researchers in the field. My goal here is to propose some future steps that can address the methodological concerns in measuring children's pragmatic development.

As Pouscoulous & Noveck (2009) and Katsos (2014) have suggested, the central issue is "the rate" at which children and adults manifest pragmatic reasoning in the experimental setting. No one doubts children's capacity to perform such computations. At issue is the extent to which children and adults compute specific implicatures. The claims are that children perform such computations less often than adults; or that children do not perform such computations where adults normally do. In the previous section, I discussed some factors that might account for these differences including processing demands, the structure of the lexicon, tolerance, as well as issues of measuring adults and children's comprehension. As Katsos (2014) pointed out, it seems reasonable to assume that all these factors play some part here. What matters is the degree to which each contributes to the outcome.

Figure 19 shows the factors that affect pragmatic computations as well as the observations of the rate of pragmatic computations in an experiment. First it is important to distinguish between factors that affect pragmatic computations and those that affect the observed rate in an experimental setting. As I showed in Study 1, given the number of

alternatives in the forced choice task (2AFC vs. 3AFC), I may get different estimates of adults' infelicity judgments, yet it is unreasonable to assume that there is a difference in adults' pragmatic capacities in these two tasks. A similar situation exists when I compare children's forced choice measures of infelicity and their open-ended feedback. In disjunctive trials where both disjuncts are true, the forced choice tasks show no sign of children detecting infelicity while the open ended responses show that children are sensitive to the infelicity of disjunction when a conjunction would have been more appropriate.

Conclusion

To conclude, I have shown that children and adults do not differ substantially in their **semantic** knowledge of the logical connectives *and* and *or*. The results were highly consistent with the current accounts that posit the semantics of *and* as logical conjunction and *or* as logical (inclusive) disjunction. With respect to pragmatic knowledge, the three-alternative forced choice judgment task showed that adults are sensitive to the infelicity of disjunctive statements when both disjuncts are true. I also showed that while the three-alternative judgment task failed to register such a sensitivity for children, my systematic analysis of children's verbal open-ended verbal feedback showed that children are sensitive to pragmatic infelicities and can provide appropriate corrections to infelicitous utterances containing logical connectives.

References

Aloni, M. (2016). Disjunction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/disjunction/

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*(1), 46–70.

Crain, S. (2012). *The emergence of meaning*. Cambridge: Cambridge University Press.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Gutzmann, D. (2014). Semantics vs. pragmatics. In L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The companion to semantics*. Oxford: Wiley.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2017). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*.

Jasbi, M., Waldon, B., & Degen, J. (submitted). *Linking hypothesis and number of response options modulate inferred scalar implicature rate*.

Katsos, N. (2014). Scalar implicature. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (Vol. 10, p. 183—198). Amsterdam: John Benjamins.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Noveck, I. A. (2001). When children are more logical than adults: Experimental

915 investigations of scalar implicature. *Cognition*, *78*(2), 165–188.
916 Pouscoulous, N., & Noveck, I. A. (2009). Going beyond semantics: The development of
917 pragmatic enrichment. In S. Foster-Cohen (Ed.), *Language acquisition* (pp. 196–215).
918 Berlin: Springer.
919 Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental
920 investigation of processing costs in implicature production. *Language Acquisition*,
921 *14*(4), 347–375.
922 Reinhart, T. (2004). The processing cost of reference set computation: Acquisition of stress
923 shift and focus. *Language Acquisition*, *12*(2), 109–155.
924 Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., & Fox, D. (2016). Children interpret
925 disjunction as conjunction: Consequences for theories of implicature and child
926 development. *Natural Language Semantics*, *24*(4), 305–352.
927 Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool
928 children. *Language Learning and Development*, *11*(2), 176–190.
929 Tieu, L., Yatsushiro, K., Cremers, A., Romoli, J., Sauerland, U., & Chemla, E. (2016). On
930 the role of alternatives in the acquisition of simple and complex disjunctions in french
931 and japanese. *Journal of Semantics*.

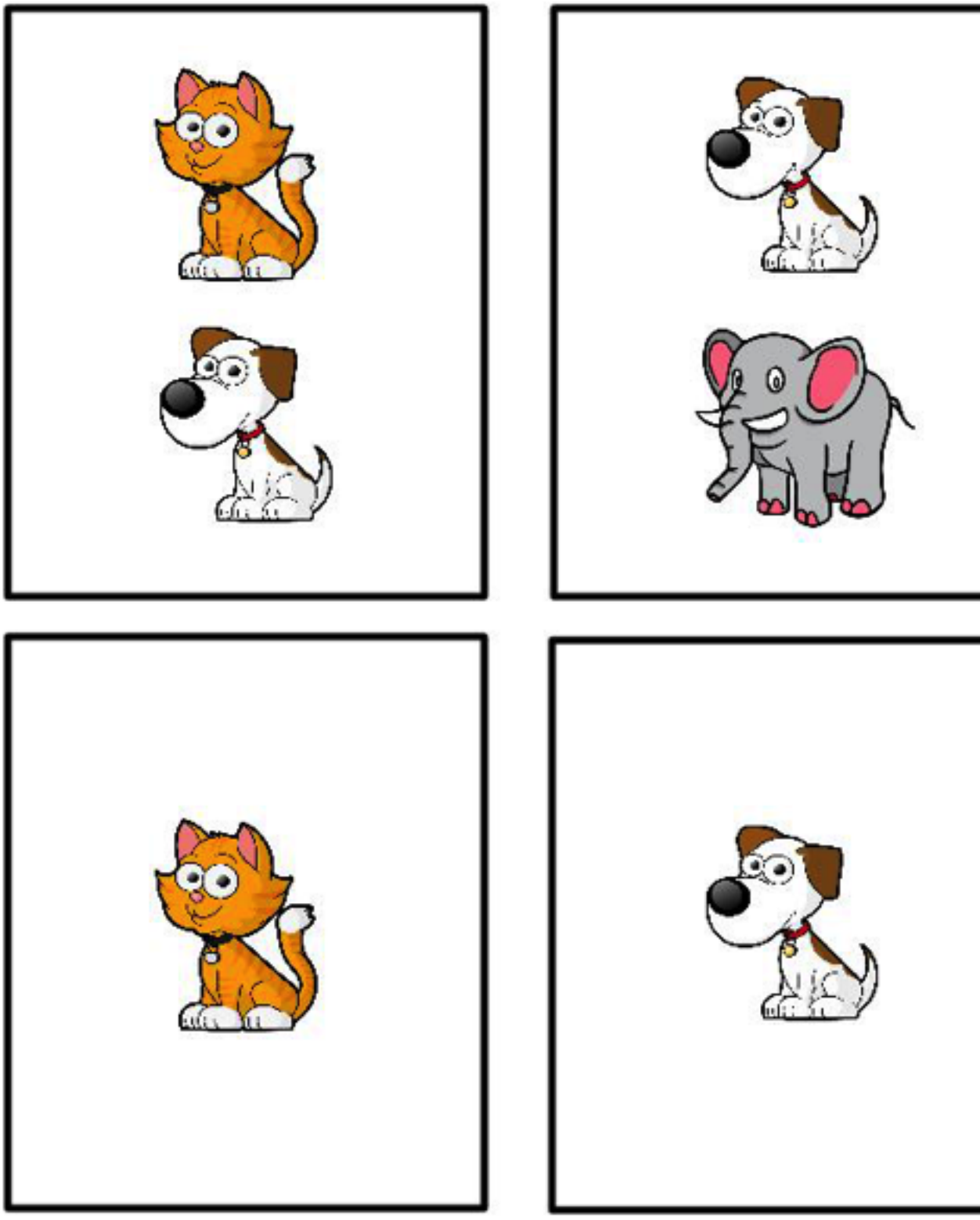
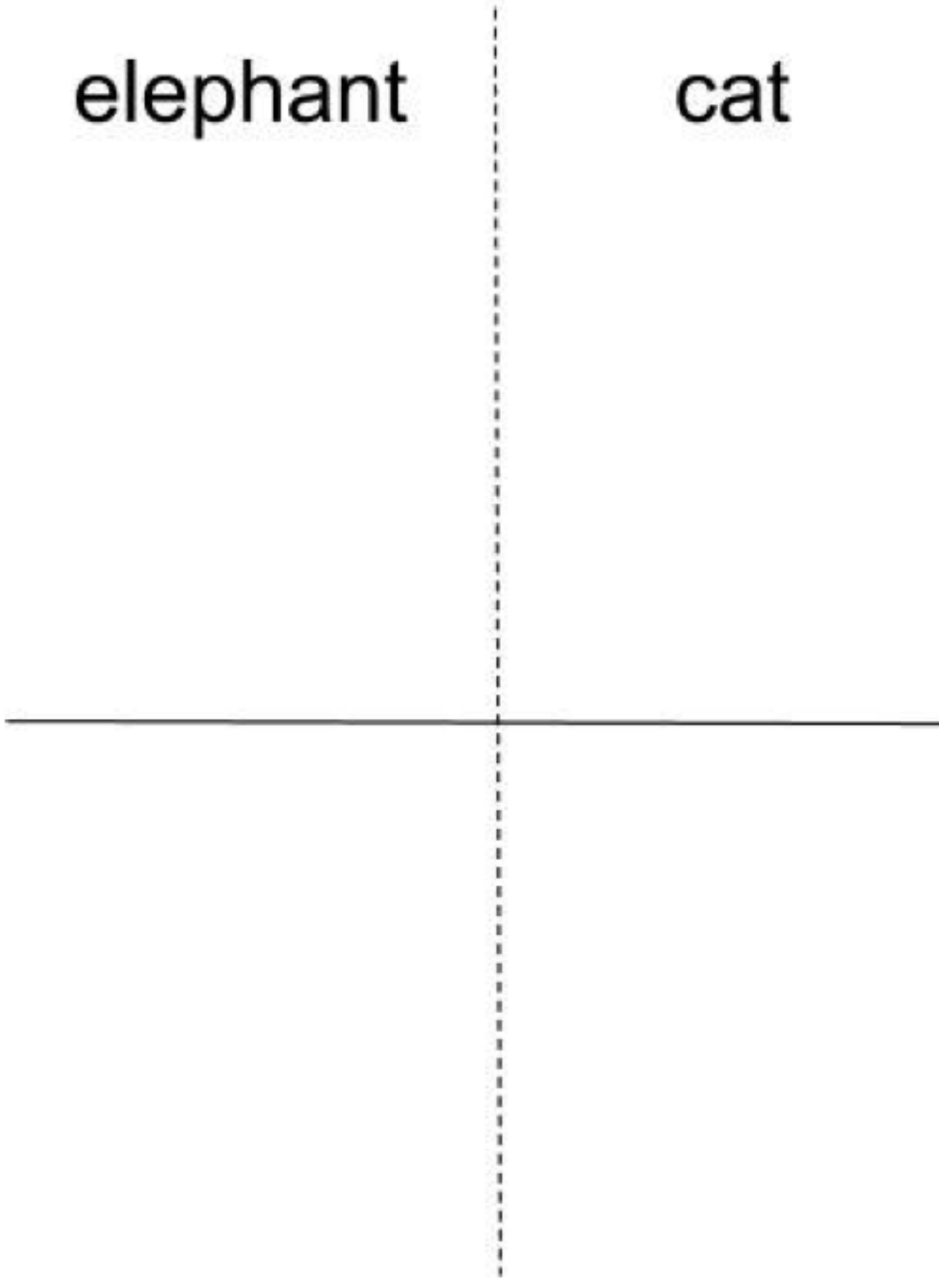
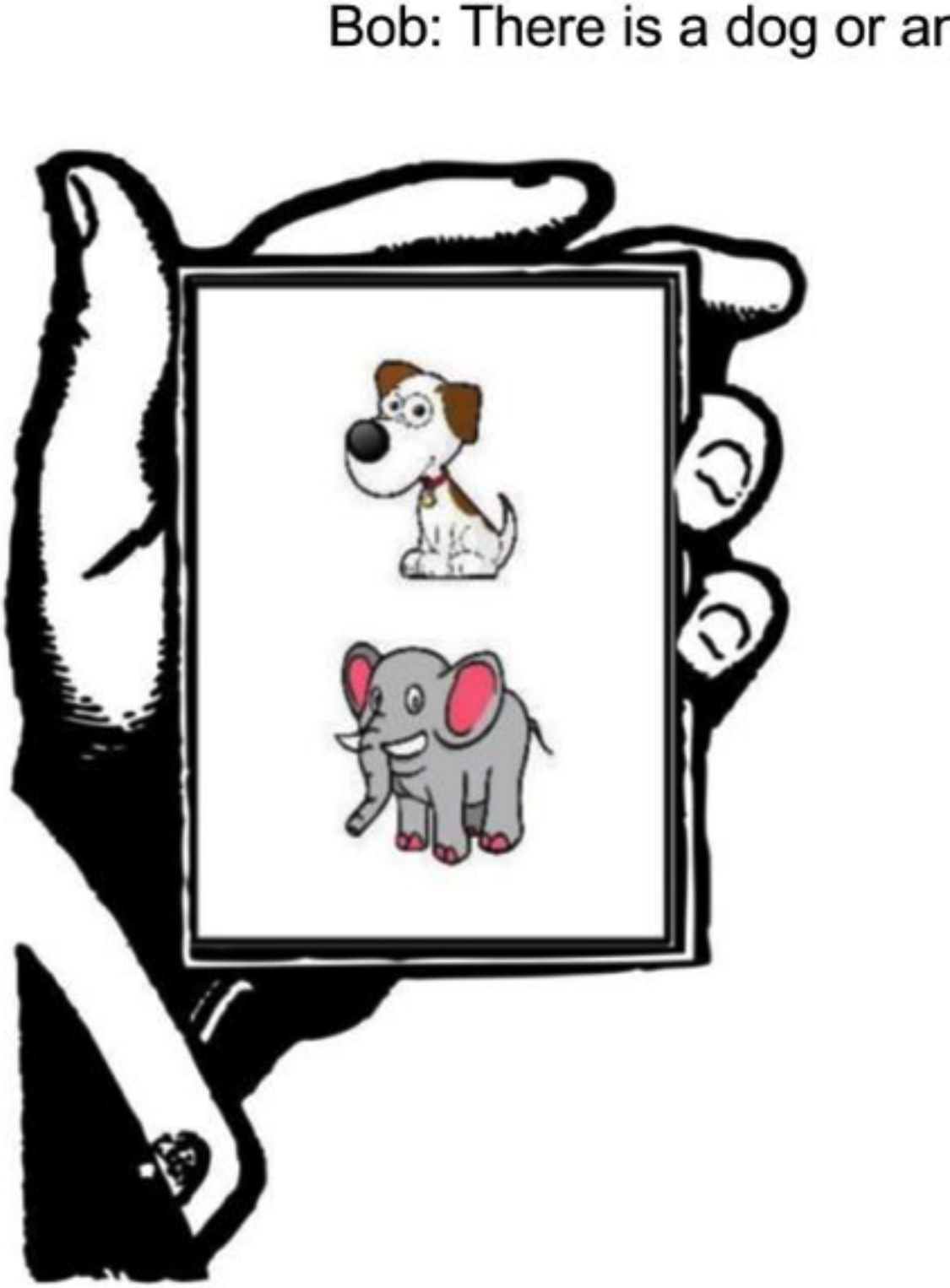
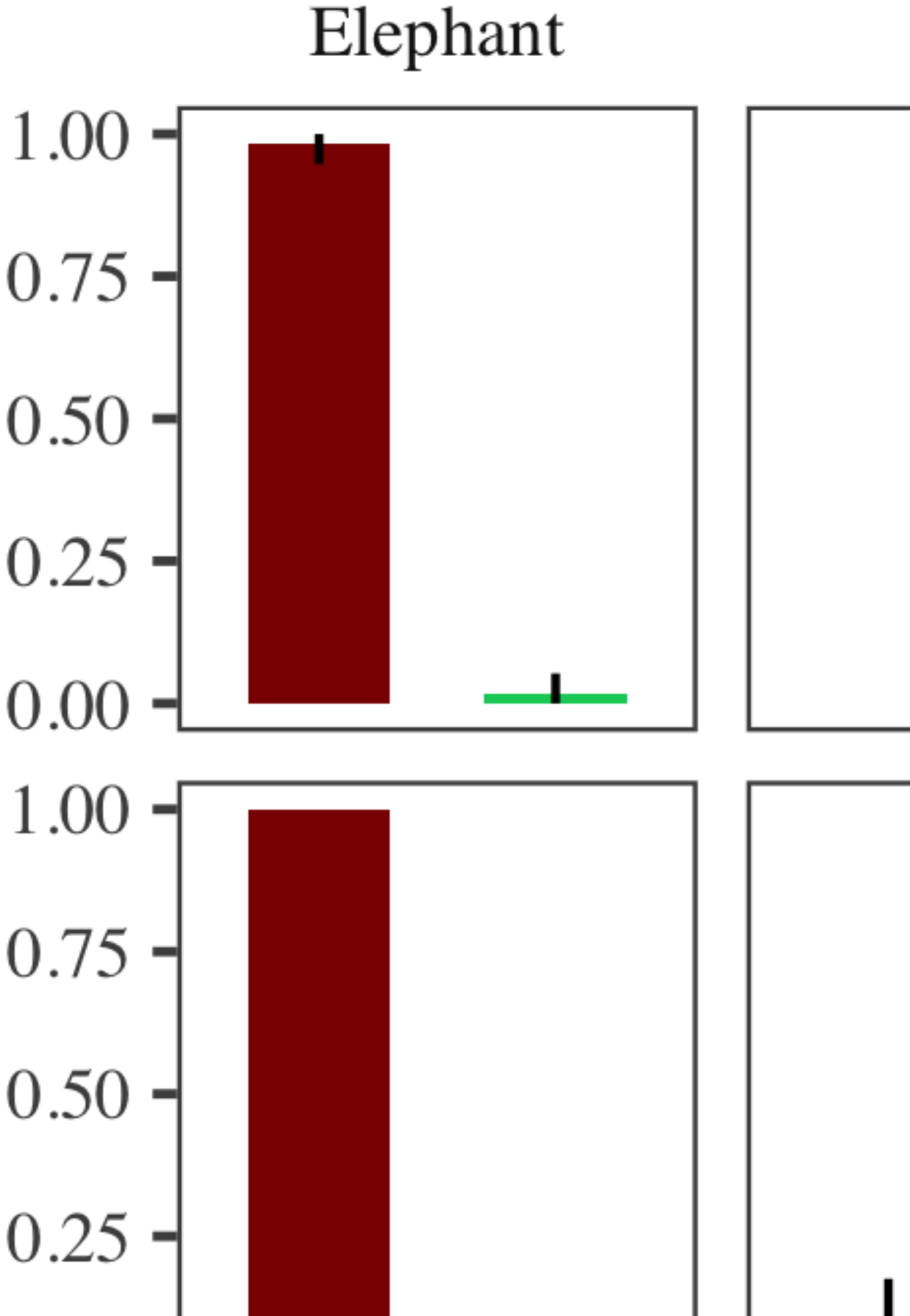
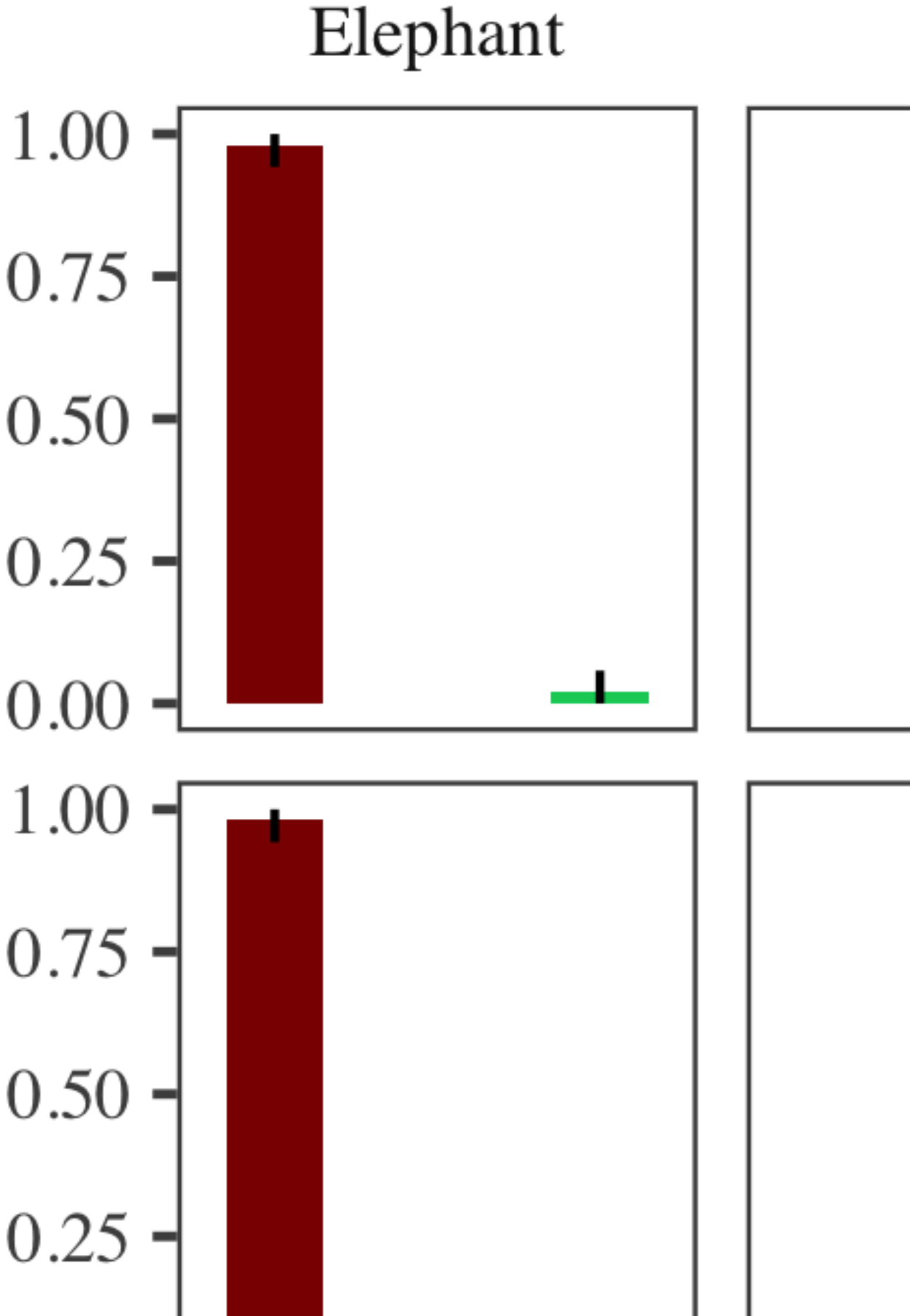


Figure 1. Cards used in the connective guessing game.


elephant
cat


Bob: There is a dog or ar


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25

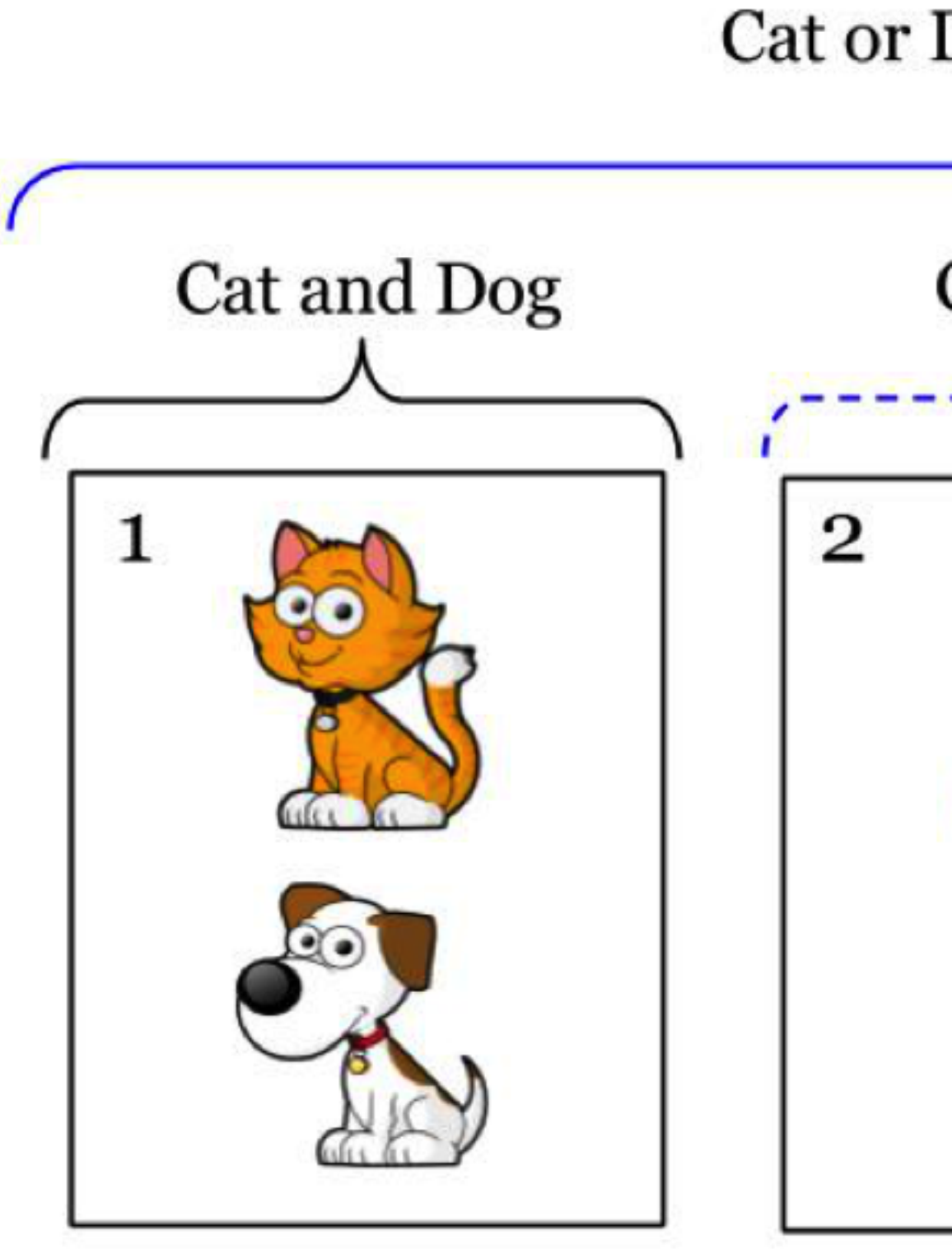


Figure 6. Example of cards referred to by a conjunction, inclusive disjunction, and exclusive disjunction.

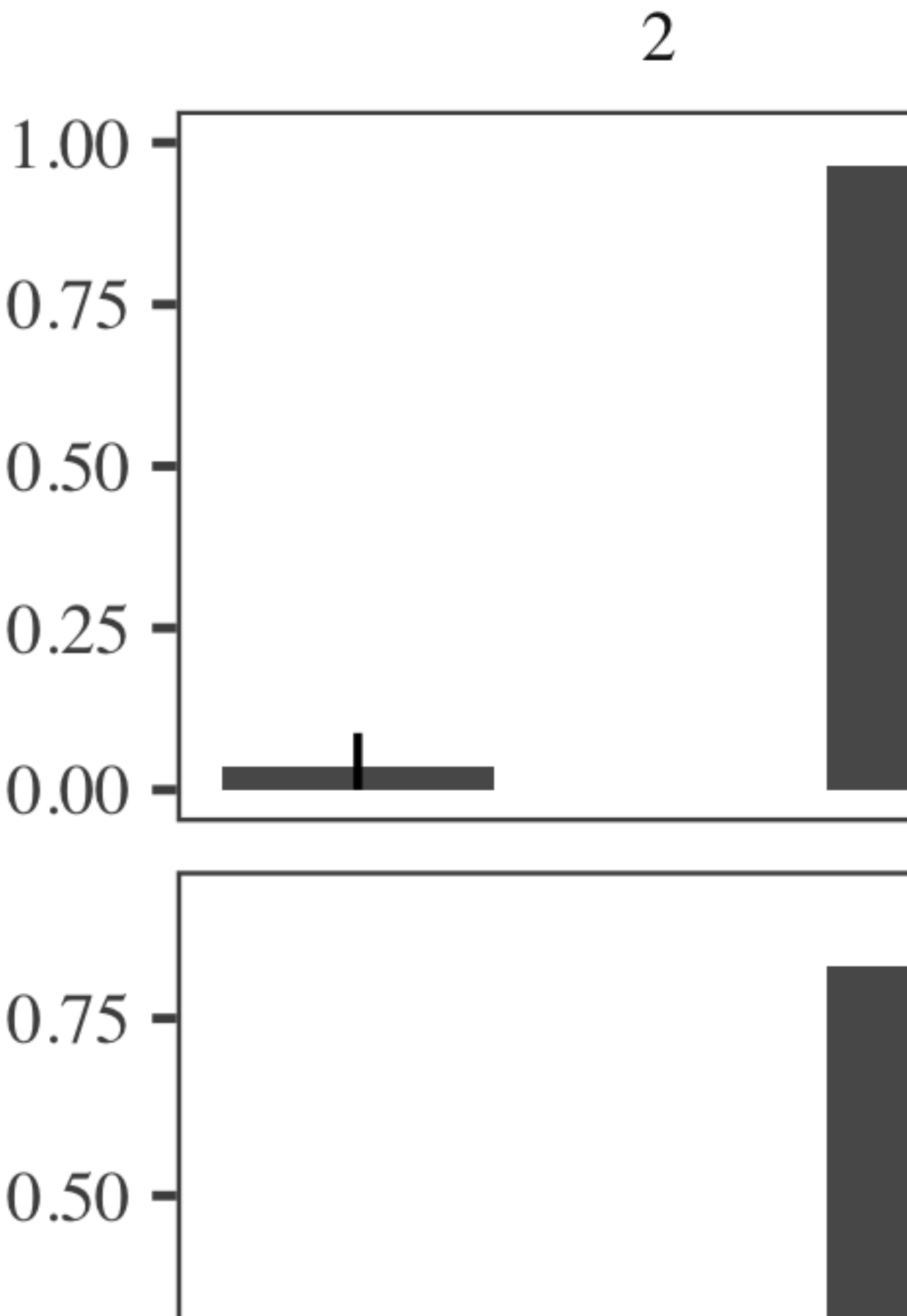
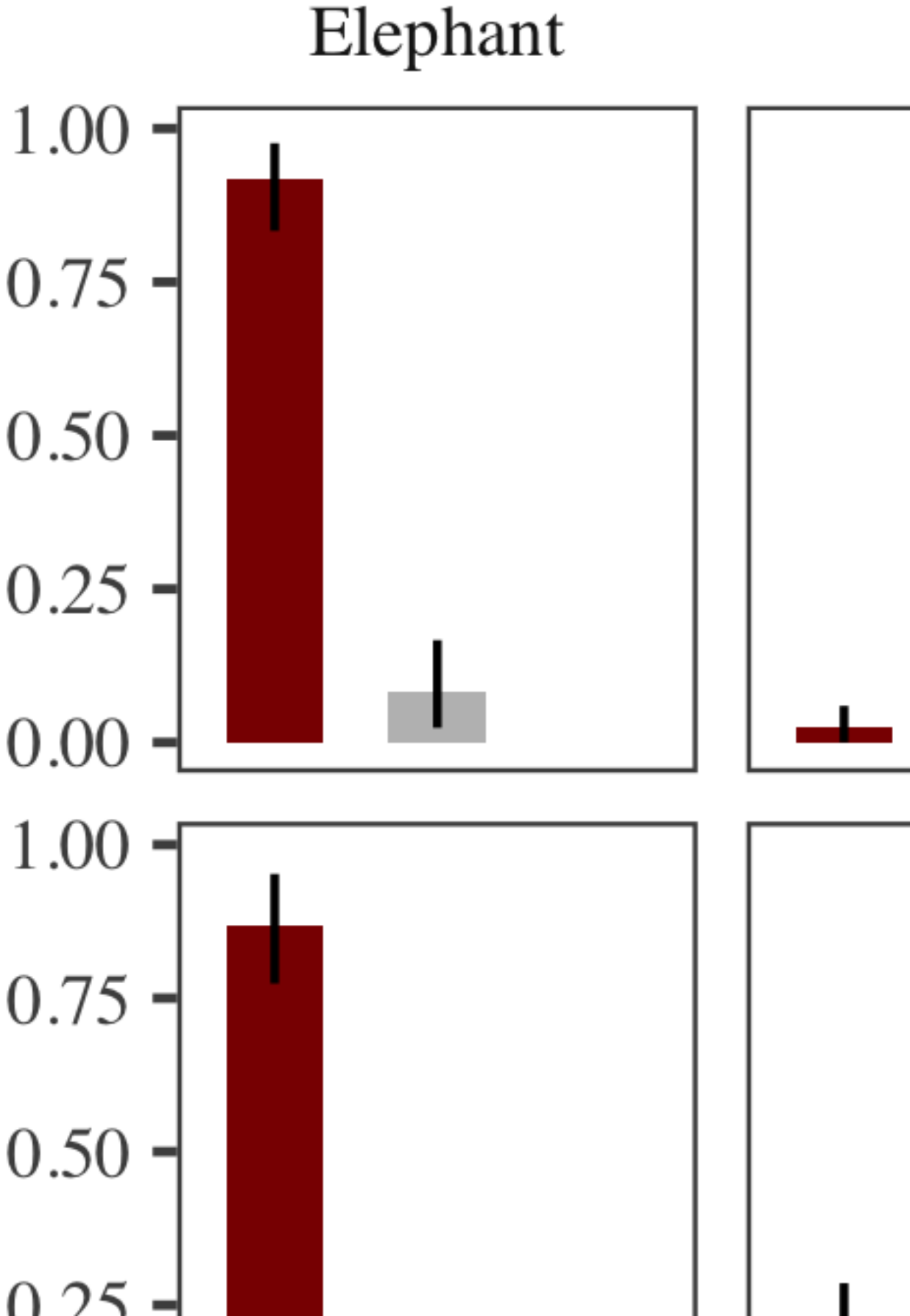
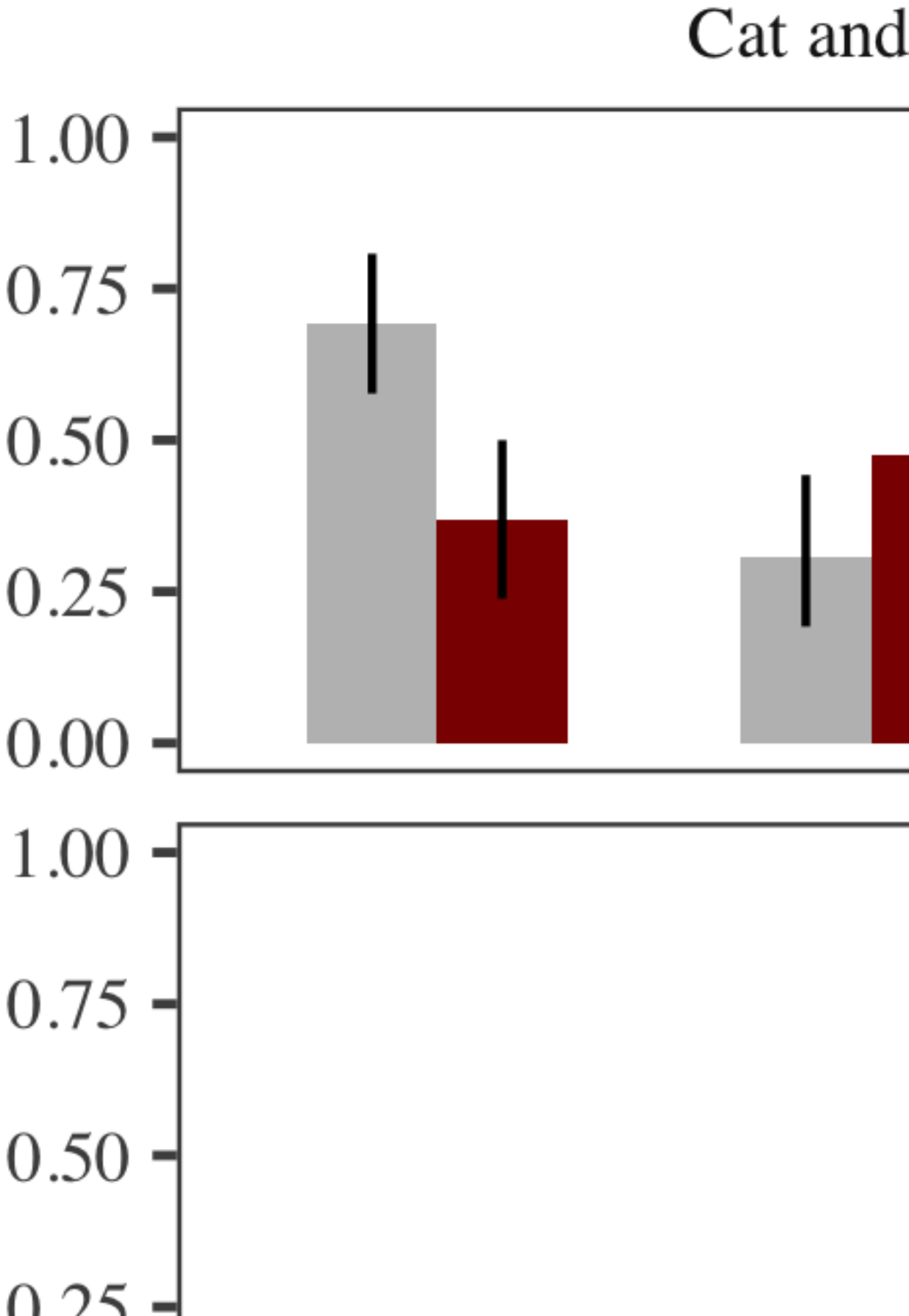

2
1.00
0.75
0.50
0.25
0.00
0.75
0.50

Figure 8. The puppet, Jazzy, with and without the sleeping mask.


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25


Cat and
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25

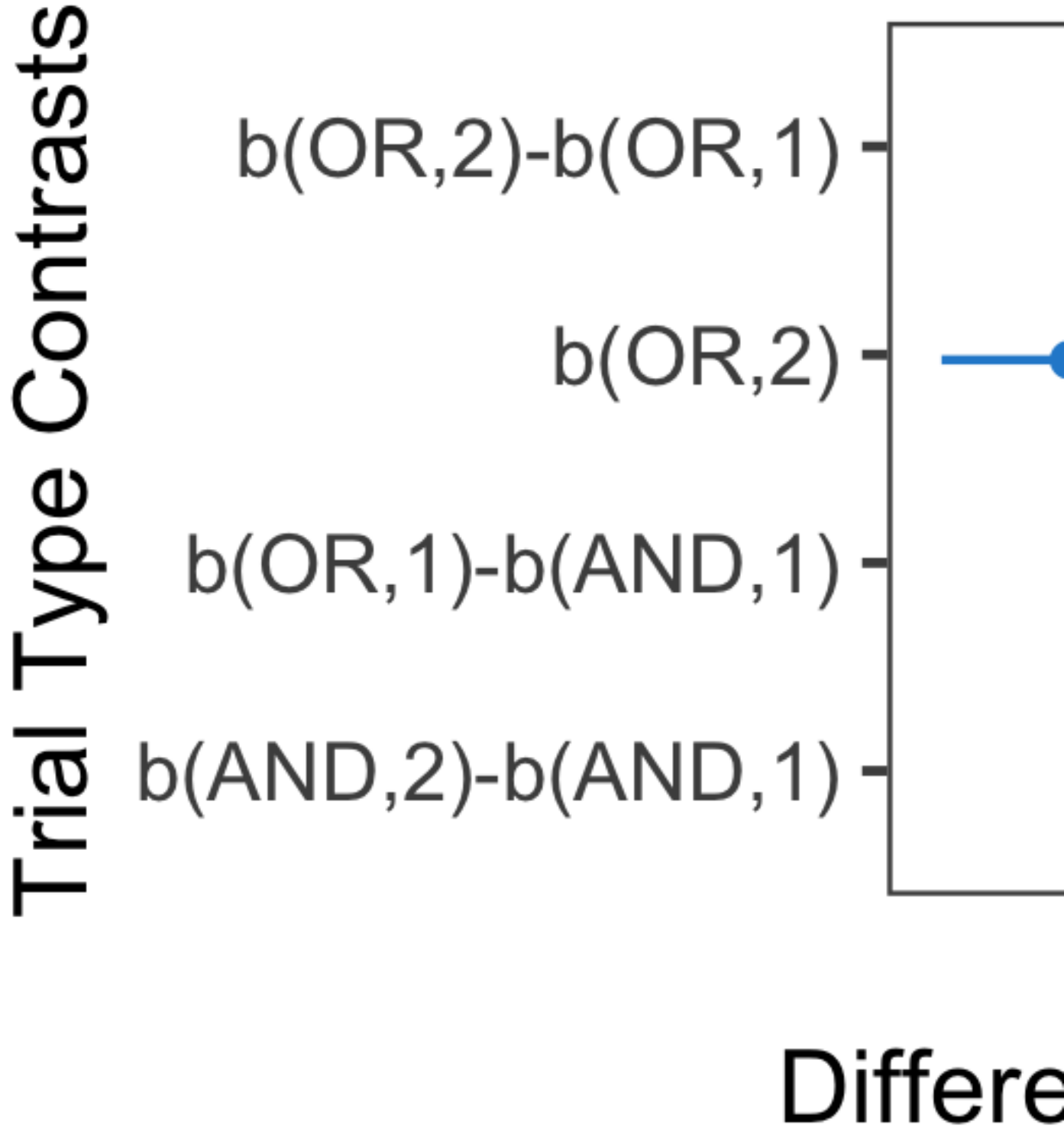
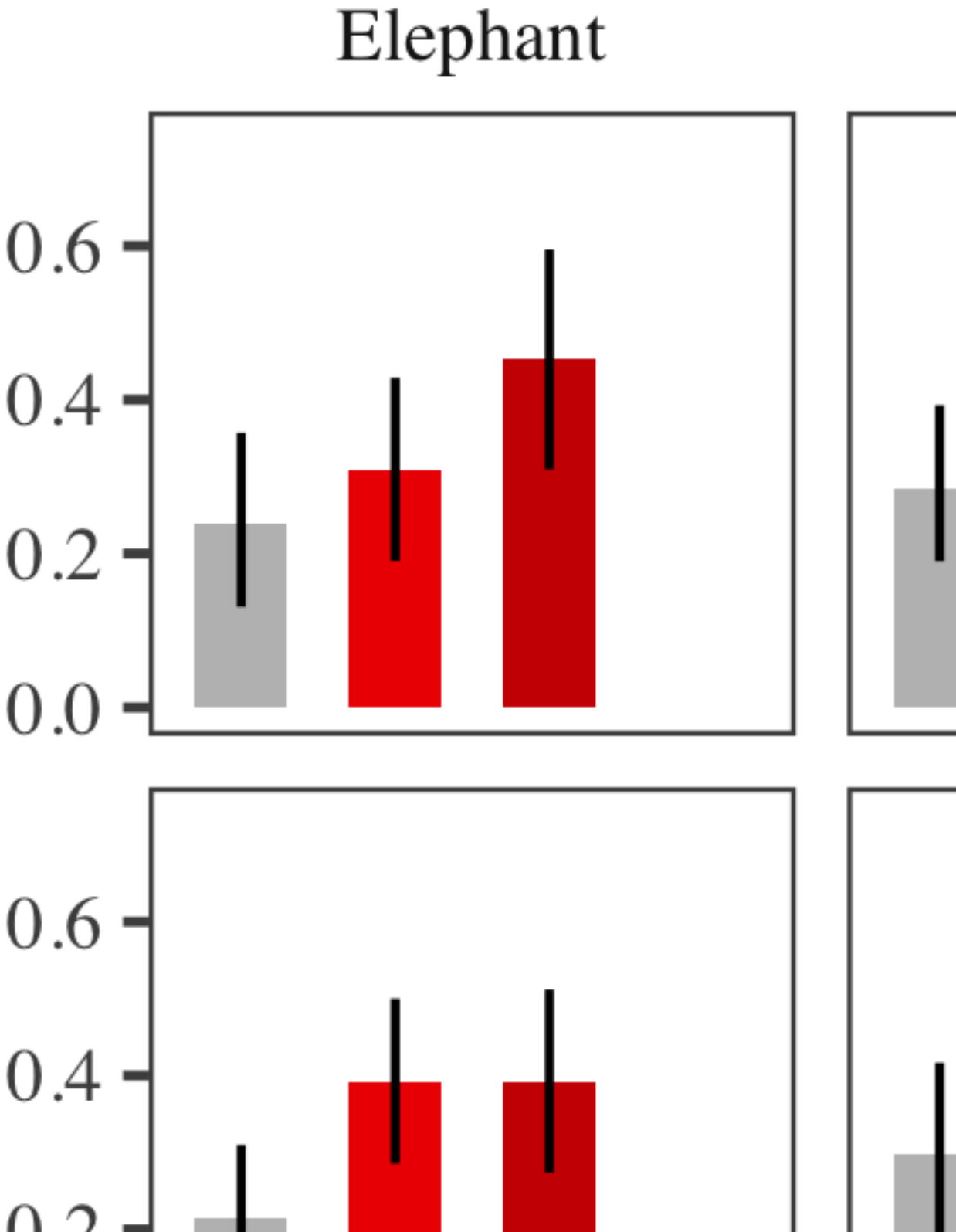
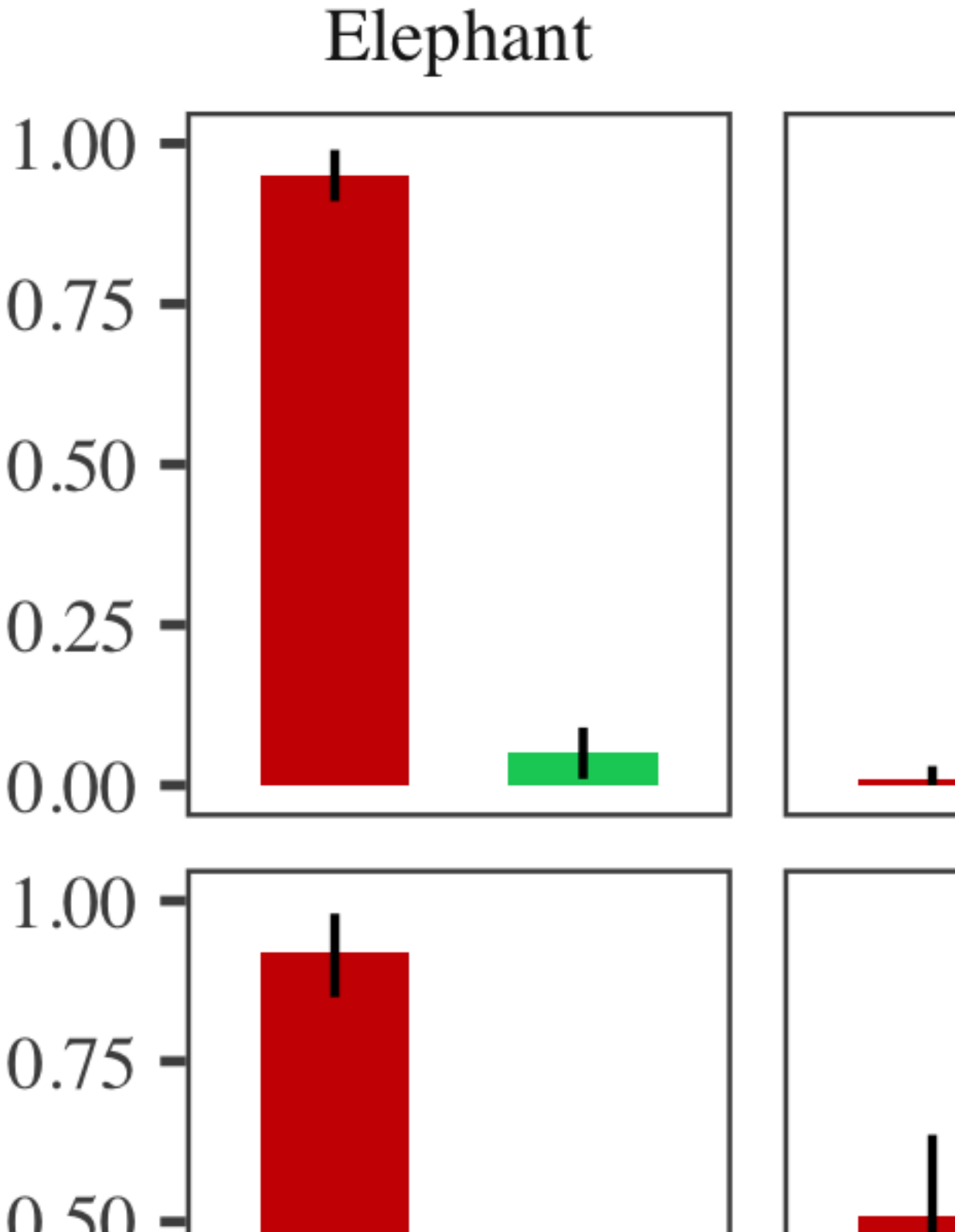
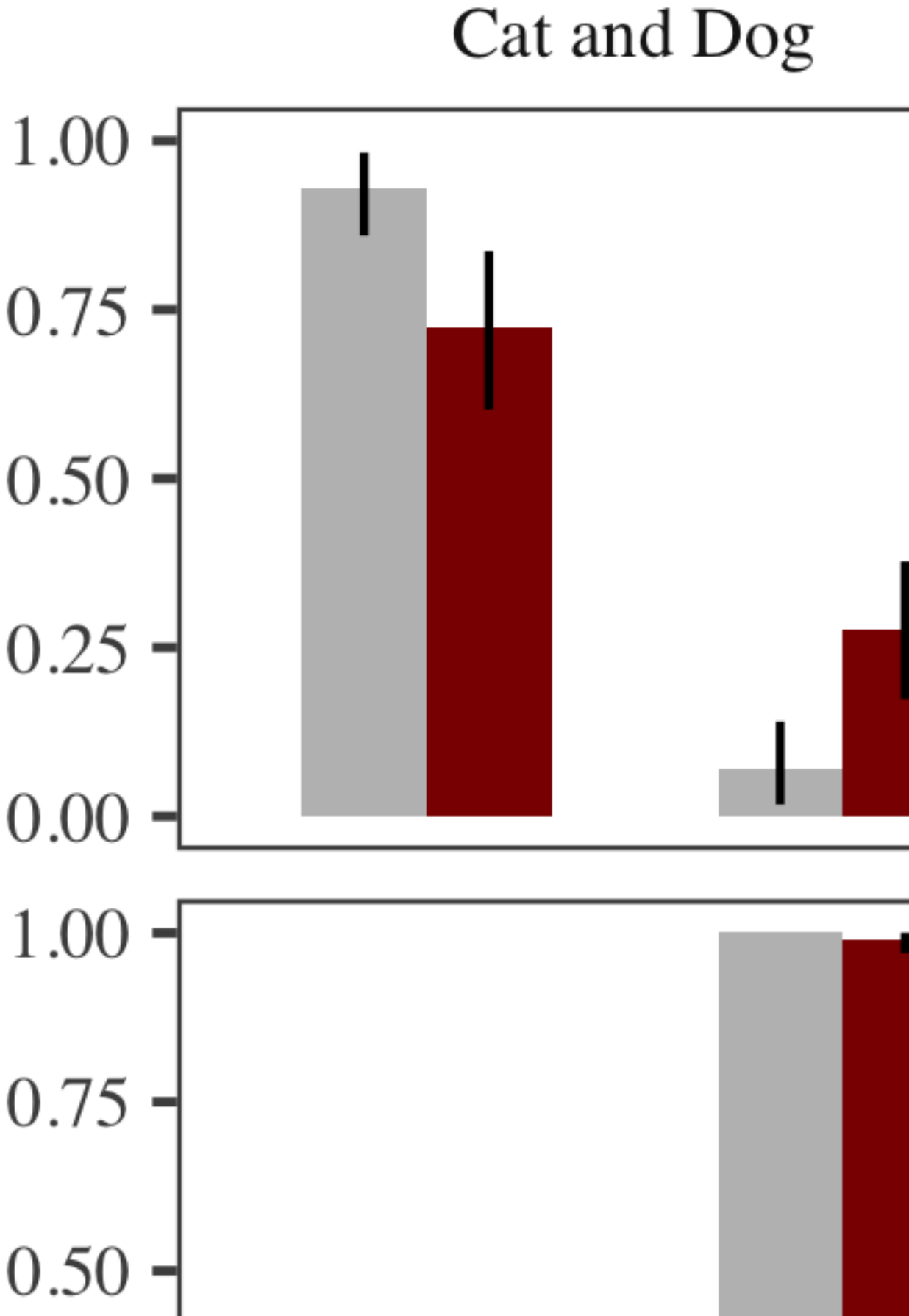
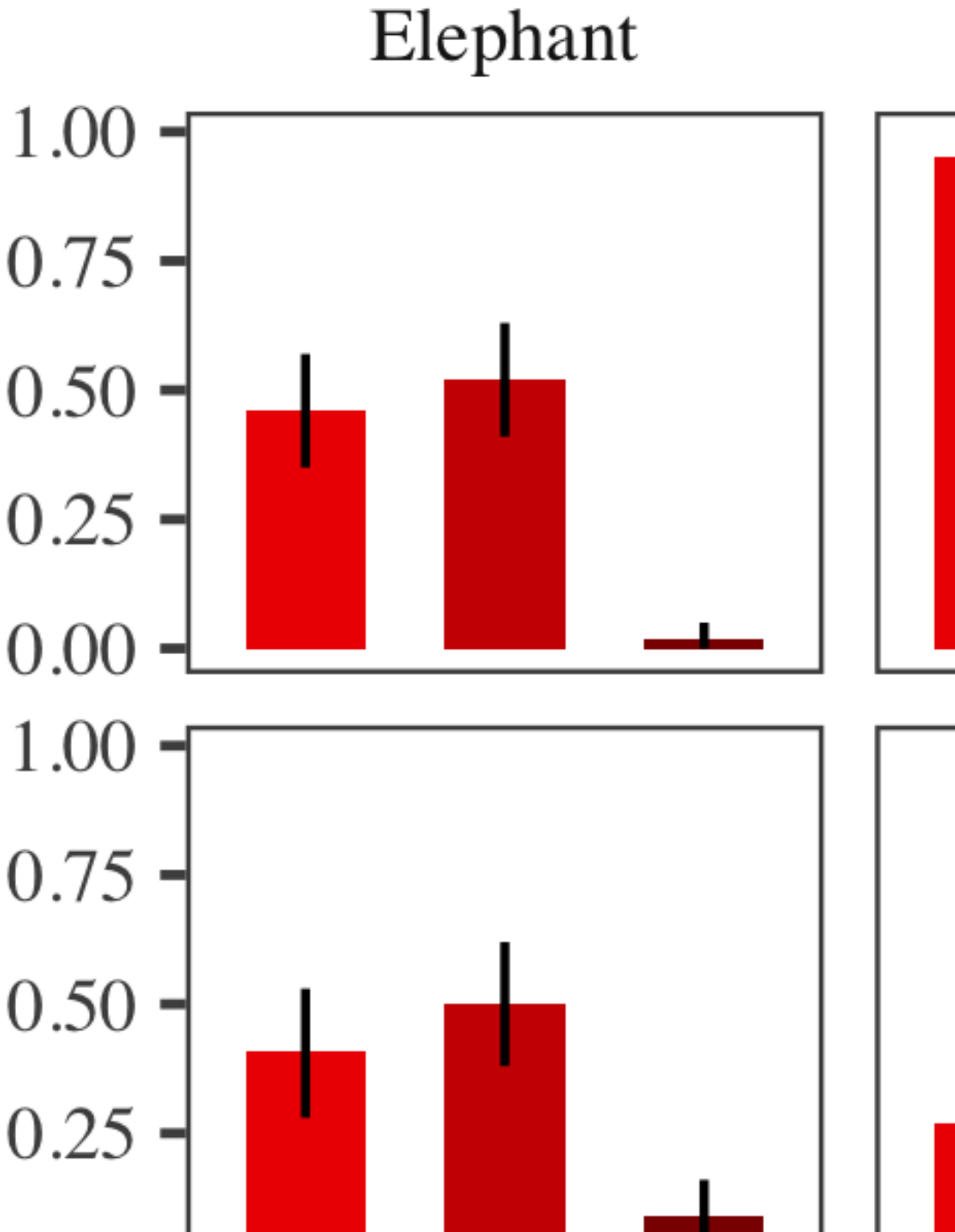


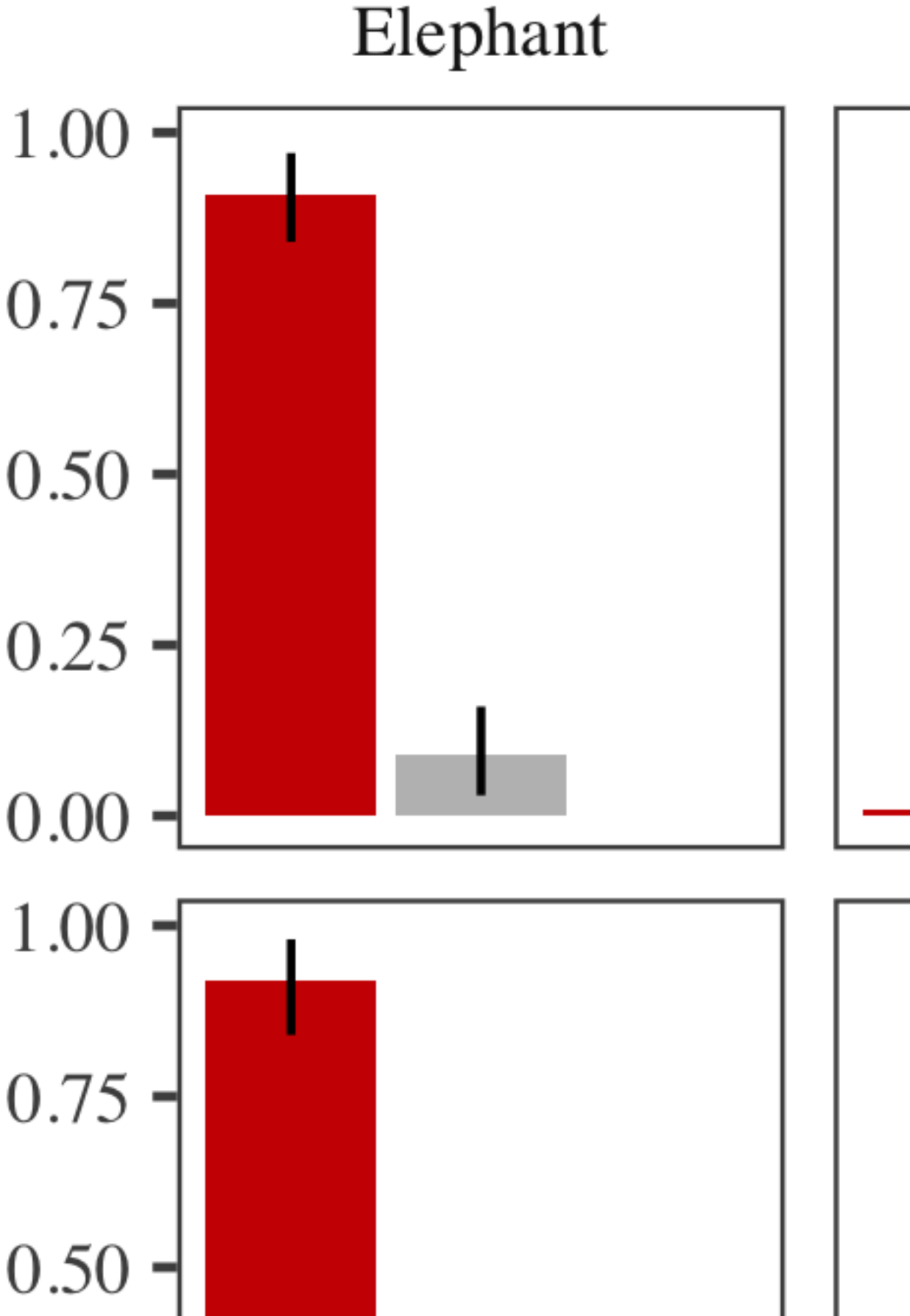
Figure 11. Coefficients capturing the relevant comparisons across conditions in ternary judgments in Study 1 and 2. In naming the coefficients like b(OR,2), OR/AND represents the connective used and the number 1/2 represents the number of animals on the card. Error bars represent 99% regions of highest posterior density.

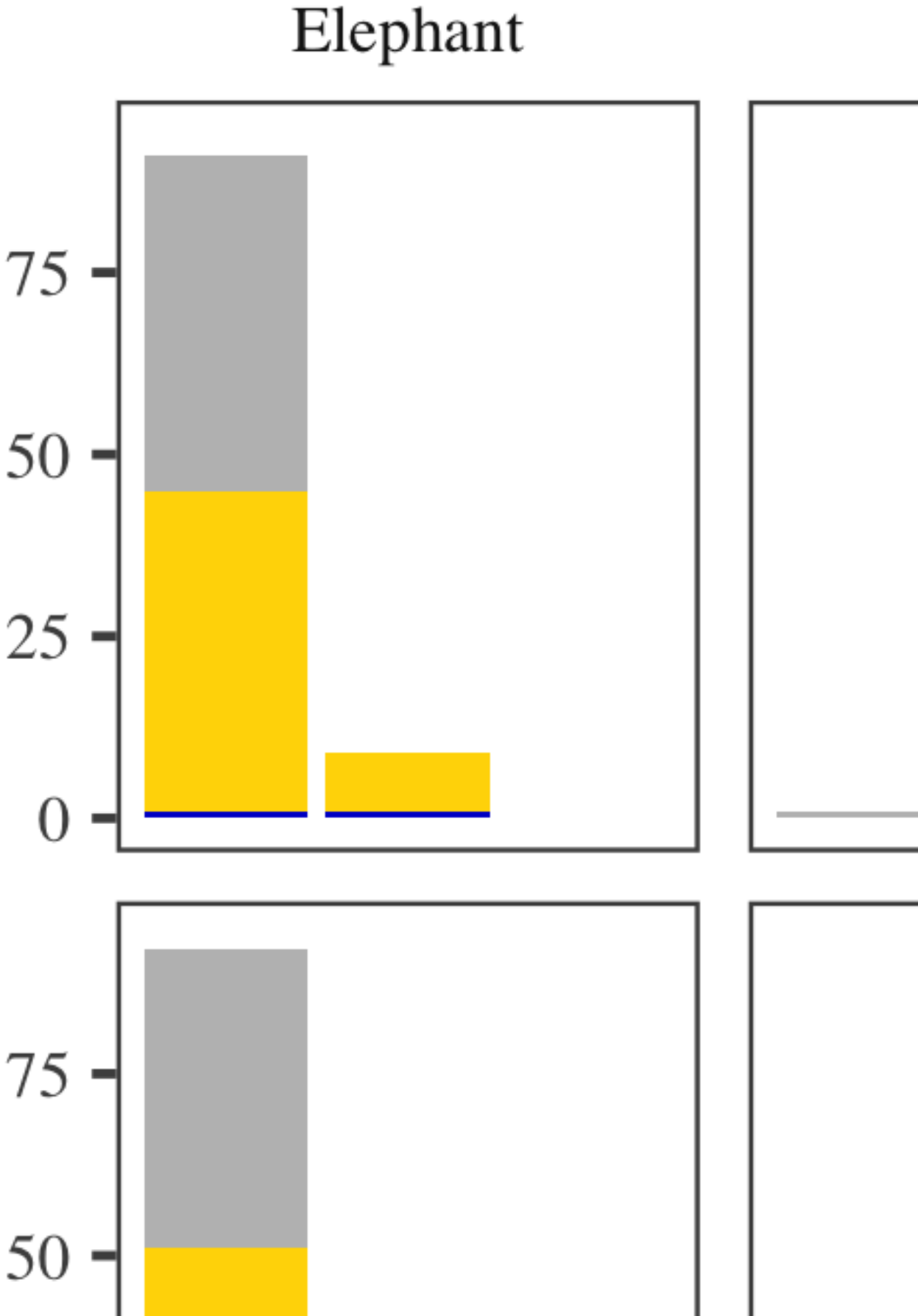

Elephant
0.6
0.4
0.2
0.0
0.6
0.4

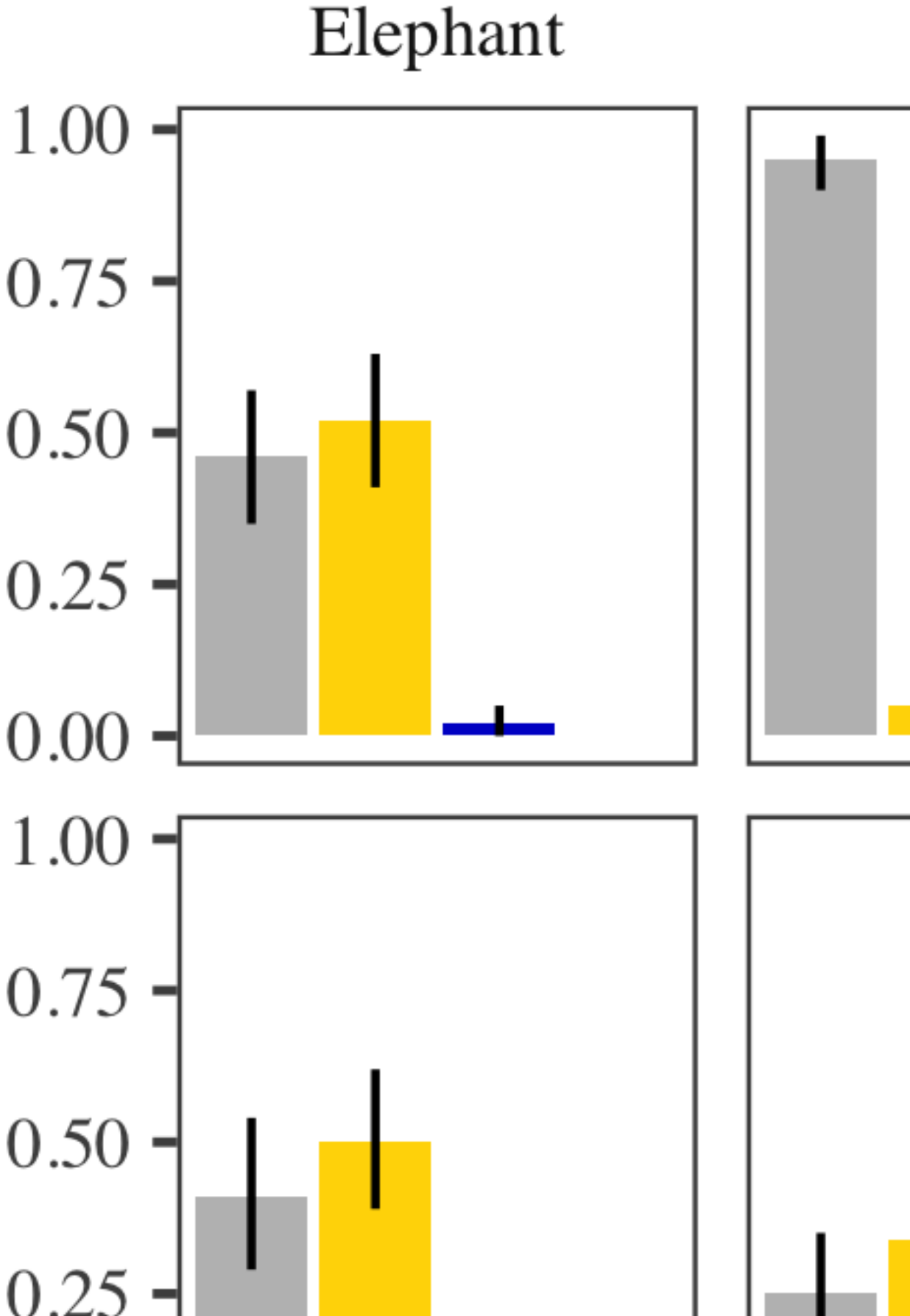

Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50

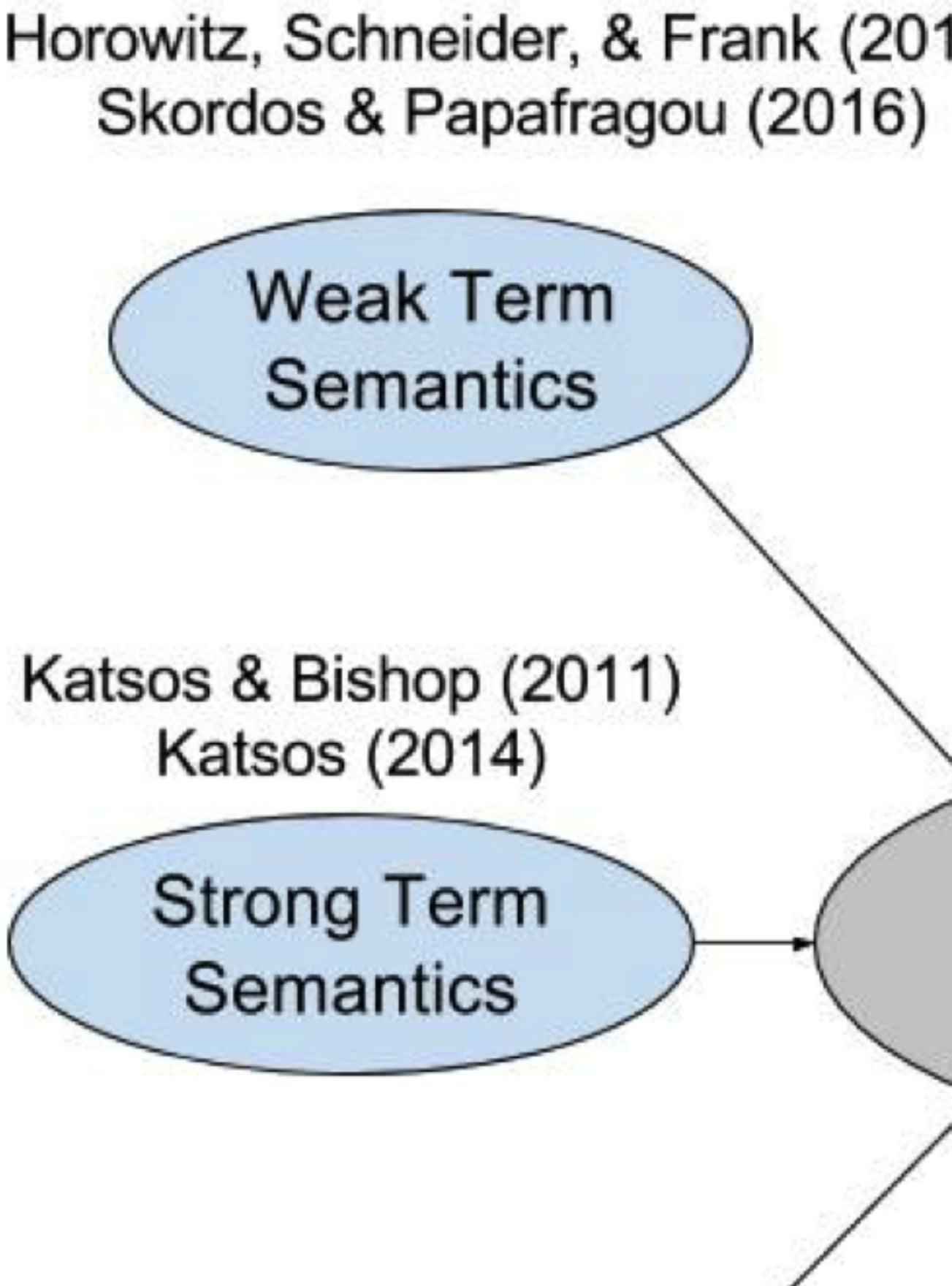

Cat and Dog
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50


Elephant
75
50
25
0
75
50


Elephant
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25


Horowitz, Schneider, & Frank (201
Skordos & Papafragou (2016)
Weak Term Semantics
Katsos & Bishop (2011)
Katsos (2014)
Strong Term Semantics