An experimental investigation of the homogeneity of conjunctions[1]

Jia REN — *University of Massachusetts Amherst*

Abstract. Nominal conjunctions in English are associated with the *all or nothing* homogeneity inference. There are two main approaches to explain the interpretations: the first derives homogeneity as implicatures (Magri, 2014; Bar-Lev and Fox, 2020), the second provides a trivalent account (Schwarzschild, 1994; Križ, 2015; Križ and Spector, 2021). The two accounts differ in their predictions on the existence of truth value gaps and the status of positive and negative sentences in non-uniform contexts. We report two experiments investigating adults' interpretation of nominal conjunctions in English. The experiments reveal that participants provide gappy judgments for sentences with conjunctions, and the gaps for positive and negative sentences in non-uniform contexts are symmetric. We discuss how the findings of the two experiments are expected on a trivalent approach to homogeneity and the open challenges they pose for the implicature approach.

Keywords: Conjunctions, Homogeneity, Plurals, Truth Value Judgment Task.

1. Introduction

1.1. The homogeneity of conjunctions

English conjunctions have *all or nothing* readings (Schwarzschild, 1994; Löbner, 2000; Beck, 2001; Szabolcsi and Haddican, 2004; Magri, 2014; Sbardolini, 2023). We refer to the *all or nothing* interpretation of conjunctions as the *homogeneity of conjunctions*. In assertive sentences, conjunctions have a universal interpretation, as exemplified in the following.

(1) Kate found Max and Tom.
$[\![(1)]\!] = found(m)(k) \wedge found(t)(k)$

The puzzle arises when conjunctions as in (1) appear under negation, as in (3). We would expect the negative sentence with conjunctions to produce a *not all* reading, as in (2). Instead, the sentence interprets as Kate found neither, as in (3).

(2) Kate didn't find both Max and Tom.

(3) Kate didn't find Max and Tom.[2]
$[\![(3)]\!] = \neg found(m)(k) \wedge \neg found(t)(k)$

This is not an isolated fact. Definite plurals also have the homogeneity interpretation. Many studies have been conducted on the homogeneity of definite plurals. (Schwarzschild, 1994, 1996; Magri, 2014; Križ, 2015; Križ and Chemla, 2015; Bar-Lev, 2021; Križ and Spector, 2021; Guerrini and Wahbe, 2023). Few studies focus on the homogeneity of nominal conjunctions. Here, we aim to fill this gap.

[1]I thank the three anonymous SuB reviewers. I thank participants of the UMass Semantics Workshop. I thank Vincent Homer, Guillermo Del Pinal. I thank Viola Schmitt, Jacopo Romoli, Lyn Tieu, Brian Dillon, Shota Momma (Negishi), Andrea Matticchio, Maik Thalmann, Yizhen Jiang, Hening Wang, Suet-Ying Lam and Mari Kugemoto.

[2]The conjunction *and* is unstressed in these two examples.

Two approaches have been proposed in the formal semantics literature. The first is an implicature treatment of the homogeneity of conjunctions, implemented in Magri (2014). The second approach accounts for the different interpretation of conjunctions by appealing to a third truth value (Schwarzschild, 1994). These different approaches can account for the interpretation of nominal conjunctions across upward- and downward-entailing linguistic environments. They differ, however, in the existence of truth value gaps and the status of positive and negative conjunction sentences in non-uniform contexts.

In this paper, we report two experiments on the interpretation of English nominal conjunctions. The experiments reveal that participants show gappy judgments for sentences with conjunctions in non-uniform contexts, and the gaps reported for positive and negative sentences are symmetric. The findings are expected in a trivalent approach to homogeneity and challenge the implicature approach. The paper is organized as follows. In the remainder of this section, we sketch the implicature and trivalent accounts and summarize their predictions. In Section 2, we review previous experimental studies on the relevant topics. Our experiments are presented in Section 3 and Section 4. In Section 5, we interpret the results in light of the theoretical predictions. Section 6 concludes the paper.

1.2. The implicature approach

Magri (2014) provides a unified implicature account to the homogeneity of definite plurals and nominal conjunctions. According to the proposal, natural language conjunctions have the semantics of the logical conjunction $\wedge$ (Keenan and Faltz, 1984; Gazdar, 1980; Partee and Rooth, 1983). When the nominal conjunction is in upward-entailing environments, one gets the strong universal interpretations without further complication. The conjunction is true if and only if both conjuncts are true. In all other conditions, the conjunction is false. Example (1), repeated below as (4), is predicted to be true if and only if Kate found both Max and Tom.

(4) Kate found Max and Tom.

Negated logical conjunctions are predicted to receive an interpretation that is weaker than what we observe in natural language. Under this analysis, example (5) is true only if Kate did not find both Max and Tom, and false if she found at least one of them, as formalized in (6).

(5) Kate didn't find Max and Tom.

(6) $\neg[\![(4)]\!] = \neg[found(m)(k) \wedge found(t)(k)]$
$= \neg found(m)(k) \vee \neg found(t)(k)$

(5) interprets as Kate found neither Max nor Tom. According to Magri (2014), the interpretation comes from exhaustification. In the proposal, there are two *and* in English, the stressed *and* and the unstressed *and*. They are Horn scale-mates of each other, and they are both interpreted as the logical conjunction $\wedge$. The only difference between them is that the stressed *and*, marked from now on as and_F, is Horn scale-mates with disjunction *or*. The scale-mate relation is not transitive in Magri (2014), therefore, *and* and *or* are not scale mates of each other. The unstressed *and* is not an alternative to the disjunction *or*. Under downward entailing environments, obligatory double strengthening takes place, thus leading to the intended interpretation. (7) interprets as Kate found neither Max nor Tom.

(7) Kate didn't find Max and Tom.
$[\![(7)]\!] = Exh[Exh[\neg find(m\ and\ t)(k)]]$
$= Exh[\neg find(m\ and\ t)(k)] \wedge \neg Exh[\neg find(m\ and_F\ t)(k)]$
$= Exh[\neg find(m\ and\ t)(k)] \wedge \neg[\neg find(m\ and_F\ t)(k) \wedge \neg\neg find(m\ or\ t)(k)]$
$= \neg find(m\ or\ t)(k)$

In summary, the truth condition of nominal conjunctions following the Boolean approach is predicted to be as follows.

	Kate found Max and Tom	Kate didn't find Max and Tom.
$found(m)(k), found(k)(t)$	True	False
$\neg found(m)(k), found(k)(t)$	False	False (Implicature violation)
$found(m)(k), \neg found(k)(t)$	False	False (Implicature violation)
$\neg found(m)(k), \neg found(k)(t)$	False	True

Table 1: Predictions: the implicature account

1.3. The trivalent approach

Schwarzschild (1994) treats *and* as a plural-forming operator. It takes two type e individuals, and outputs an individual of the same type. The output is the mereological sum of the two input individuals.

(8) $[\![and]\!] = \lambda x_e.\lambda y_e.x \oplus y.$

Given the semantics, a conjunction denotes a plural individual with the conjuncts as its parts. Nominal conjunctions are similar to definite plurals in that both denote plural entities. This view is supported by the parallel behaviors between nominal conjunctions and other plural nouns. For example, plural definite descriptions also give rise to the homogeneity effect (Schwarzschild, 1994; Gajewski, 2005; Križ, 2016; Bar-Lev, 2021; Križ and Spector, 2021).

Schwarzschild (1994) further treats homogeneity as a presupposition in plural predications. It is proposed that the extension of certain plural predicates is closed under plural formation. Given the property and Cooper's semantics (1983) for presuppositions, the homogeneity presupposition arises. When applying a predicate P to a plurality X, it is presupposed that X is homogeneous for P such that either all parts of X are P, or none are.

(9) Kate found the Max and Tom.

In (9), the plural X is *Max and Tom*, the predicate P is *found the cats* In (9), the predicate P is applied to X. Thus, it is presupposed that either Kate found both Max and Tom, or Kate found neither of them. The predicted interpretation of (9) is thus the former.

(10) Kate didn't find the Max and Tom.

Following the same logic, (10) is interpreted as Kate found neither Max nor Tom. In a scenario where Kate found either Max or Tom, asserting (9) and (10) will lead to a presupposition failure. Thus, we make the following predictions.

	Kate found Max and Tom.	Kate didn't find Max and Tom.
$found(m)(k), found(k)(t)$	True	False
$\neg found(m)(k), found(k)(t)$	Presupposition failure	Presupposition failure
$found(m)(k), \neg found(k)(t)$	Presupposition failure	Presupposition failure
$\neg found(m)(k), \neg found(k)(t)$	False	True

Table 2: Predictions: the presupposition account

1.4. Predictions

Both approaches are designed to capture the homogeneity of conjunctions. The two approaches diverge, however, on two predictions: the existence of truth value gaps and the status of positive versus negative conjunction sentences under nonuniform contexts. We turn to each of these next.

1.4.1. The existence of gap

An important feature related to the homogeneity of definite plurals is the so-called *gappiness*. Križ and Chemla (2015) observe that under nonuniform scenarios, people give neither true nor false judgment to sentences with plural definites in ternary judgment tasks.

(11) Kate found the cats.

(11) is judged to be true if Kate found all the cats, and it is judged to be false if Kate found none of the cats. If Kate found some, but not all, cats, (11) has a high probability of getting the neither true nor false response. This judgment is experimentally confirmed in Križ and Chemla (2015).

(12) Kate found Max and Tom.

Following Magri (2014), (12) is predicted to be true if and only if Kate found Max and Tom, (12) is predicted to be false for all other scenarios. Following Schwarzschild (1994), we assume that nominal conjunctions behave like definite plurals. We thus expect (12) to have the same truth value judgement as (11). In a scenario where Kate found both Max and Tom, (12) is considered true. In a scenario where Kate found neither Max nor Tom, the account predicts that the participants will judge (12) as false. In a scenario where Kate found only Max or Kate found only Tom, presupposition failure arises.

We summarize the different predictions made by the two accounts in Table 3. Whether gappiness exists in the interpretation of nominal conjunctions has not been discussed in previous studies. The existence or non-existence of gappiness will be, however, an important support or challenge to existing proposals.

1.4.2. The status of positive vs. negative sentences

The two approaches also make different predictions regarding the status of positive and negative sentences with nominal conjunctions.

Following Magri (2014), *and* has the semantics of logical conjunctions. In upward-entailing

Kate found Max and Tom	Magri (2014)	Schwarzschild (1994)
$found(m)(k), found(k)(t)$	True	True
$\neg found(m)(k), found(k)(t)$	False	Presupposition failure
$found(m)(k), \neg found(k)(t)$	False	Presupposition failure
$\neg found(m)(k), \neg found(k)(t)$	False	False

Table 3: Divergent predictions for (12)

environments, the literal meaning of the sentence is also the strongest alternative, and thus exhaustification is vacuous. In a downward-entailing environment, the negation of the conjunction equals the disjunction of the negated conjuncts. Double exhaustification takes place and yields an implicature which is stronger than the literal meaning. Experimental studies have shown that a sentence with a true literal meaning and false implicature has a high probability of getting intermediate judgements in ternary judgement task (Katsos and Bishop, 2011; Tieu et al., 2016; Renans et al., 2018). Following Magri (2014), one thus predicts that under non-uniform contexts, the positive and the negative sentences will be judged differently in ternary judgement tasks. In positive contexts, the literal meaning of the conjunction gives the *both* interpretation, gappiness is not predicted. Under negative contexts, when the literal meaning is true but the implicature is false, gappy judgement is predicted.

In the presupposition account, gappiness is predicted in case of presupposition failure (Abrusán and Szendröi, 2012). For sentences with nominal conjunctions in both positive and negative contexts, if the scenario is non-uniform with respect to the conjuncts, a presupposition failure arises. The implicature and presupposition approaches make different predictions for sentences under non-uniform scenarios. We can see the symmetric vs. nonsymmetric contrast by comparing Tables 3 and 4. Under nonuniform scenarios in which Kate found either Max or Tom, Magri (2014) predicts (12) to be plainly false and predicts (13) to get gappy judgements. Schwarzschild (1994) predicts (12) and (13) to obtain symmetric gappy judgments, as both have presupposition failures.

(13) Kate didn't find Max and Tom.

Kate didn't find Max and Tom	The implicature account	The presupposition account
$found(m)(k), found(k)(t)$	False	False
$\neg found(m)(k), found(k)(t)$	False (Implicature violation)	Presupposition failure
$found(m)(k), \neg found(k)(t)$	False (Implicature violation)	Presupposition failure
$\neg found(m)(k), \neg found(k)(t)$	True	True

Table 4: Divergent predictions for (13)

2. Previous studies

2.1. Using ternary judgment tasks in experimental semantics

Ternary judgment tasks are used in experimental semantics and pragmatics to investigate phenomena, including implicatures (Katsos and Bishop, 2011; Tieu et al., 2017; Renans et al.,

2018), presuppositions (Abrusán and Szendröi, 2012), the interpretation of definite plurals (Križ and Chemla, 2015; Tieu et al., 2019; Augurzky et al., 2023), etc.

In Abrusán and Szendröi (2012), it is observed that in ternary judgement tasks looking at presuppositions, presupposition failures are correlated with intermediate judgements. In Katsos and Bishop (2011), it is observed that in ternary judgement tasks looking at implicatures, participants reserve the highest valued judgement for true sentences, the intermediate judgement is reported for sentences which contain false implicature and true literal meanings, and the lowest valued judgement is reported for sentences which contain false literal meanings. Tieu et al. (2017) look at the free choice inference, and observe a similar contrast between literal false sentences and sentences with false implicatures. We thus adopt the linking hypothesis that for sentences with false implicatures or presupposition failure, there is a higher probability of intermediate judgements in ternary judgement tasks.

2.2. Križ and Chemla (2015)

Križ and Chemla (2015) develop and deploy methods to detect and characterize truth value gaps focusing on the homogeneity of definite plurals.

After reviewing the pros and cons of previous ternary judgment experiments, Križ and Chemla (2015) uses two methods to diagnose truth value gaps. The first method is to collect standard truth-value judgments for both positive and negative sentences, thus identifying a gap as a discrepancy between the two cases. Sentences like those below are presented to the participants.

(14) a. The squares are green.
b. The squares are not green.

Along with the sentence, a picture with colored squares is provided. The critical scenarios are those in which some but not all squares are of the mentioned color, for instance, green. Some participants are asked to choose between *completely true* and *not completely true*. Some other participants are asked to choose between *completely false* and *not completely false*. A gap is found if a sentence is judged as *not completely true* by the first group of participants and *not completely false* but by the second group of participants.

The other method is ternary judgment tasks. For sentences with the same contexts, participants are asked to choose between *completely true*, *neither* and *completely false*. Truth value gaps are observed in both binary and ternary judgement tasks, and the two methods show strikingly similar results. Thus, we will mainly show the result of ternary judgment task without loss in generality.

It has been observed *all*-NPs do not create truth value gaps (Brisson (1998), Križ and Spector (2021), to name a few). When the sentences are provided under scenarios which are non-uniform, definite plurals and *all*-NPs behave differently. For sentences with *all*-NPs, participants either all choose *true*, or all choose *false*, except for a few outliers. For sentences with definite plurals, the intermediate *neither completely true nor completely false* option is used by the majority of participants. Gappiness is not observed for *all*-NPs, as predicted; gappiness is observed for definite plurals, as predicted by theories on the homogeneity of definite plurals (Križ (2015), Bar-Lev (2021), to name a few).

3. Experiment 1

We designed a ternary truth value judgement task to assess the interpretations that English speakers assign to assertive sentences with nominal conjunctions. The experimental design is based on the previous experiments by Križ and Chemla on definite plurals. The goal is to investigate whether there is gappiness in the interpretation of nominal conjunctions.

3.1. Methods

3.1.1. Participants

We tested 42 native English speakers. Participants were recruited through Prolific. The participants were paid 1.5 USD for for the 8 minute study, for a pay rate of 16 USD /hour.

3.1.2. Procedures

In the experiment, participants were given a background story about a toy finding game. In each case, a child was assigned to a room with four hidden waving lucky cats. The child was encouraged to find as many cats as possible. Each cat has a unique name. The result of the toy finding game was presented, with a sentence describing the result provided under the result. The participants then had to decide upon seeing the result how right the sentence had been. Following the method in ternary judgment tasks, we provided participants with three response options: *Completely false*, *Neither completely false nor completely true*, *Completely true*. Participants indicated their response by clicking the button that had the desired response.

3.1.3. Materials

We adopted a 2×3 balanced Latin square design. The two factors are: structure (conjunctions vs. "all"-NPs) and display (true scenarios, non-uniform scenarios, false scenarios). In total, participants received two training items, followed by a semi-randomized sequence of 15 targets (5 under each display), 15 controls (5 under each display) and 30 filler sentences, for a total of 60 experimental trials.

The critical sentences are as follows.

(15) Target sentences
[Name] found [Cat name 1], [Cat name 2], [Cat name 3] and [Cat name 4].
Example: Elizabeth found Max, Coco, Luna and Tom.

(16) Control sentences
[Name] found all the cats.
Example: Nina found all the cats.

Each of the trails involved a display that contained four objects. The critical sentences are provided under three types of display. In the first kind of display, all four cats were found. In the second kind of display, two out of four cats were found, the remaining two were not found. In the third kind of display, all four cats were not found. We used a green circle around the object to indicate that the cat was found by the child, and a red circle with a line through it to indicate that the cat was not found by the child. The cat icons were counterbalanced for their positions and frequencies.

Alongside the critical sentences, we also include some filler sentences. These involved cases where the objects were mentioned using numerals, or quantifiers like *many* and *a few*.

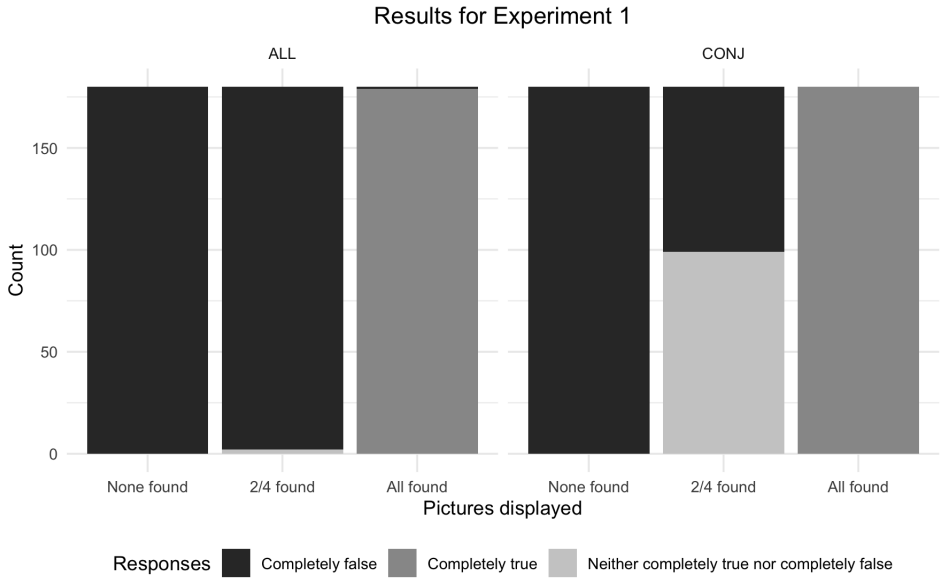
(17) Examples of Filler sentences
   a. Nina found more than two cats.
   b. Roberts found less than two cats.
   c. Mari found many cats.
   d. Kevin found a few cats.

3.2. Data analysis

Given that the response variable is discrete and ordinal, we believe that ordinal regression methods are the most appropriate. The factors we involve in the modeling include explanatory variables (display and structure), and the response variable, which has three levels, *Completely true*, *Neither completely true nor completely false*, *Completely false*. We fit the data to the Cumulative Link Mixed Model using the clmm function in the ordinal package in R. (Christensen, 2019) Previous experiments involving ternary judgment tasks use the same regression analysis (Tieu et al., 2017; Augurzky et al., 2023). We used contrast codings. For *Display*, there are three levels, i.e., 0/4 cats are found, 2/4 cats are found, 4/4 cats are found. We treat the non-uniform display, 2/4 as the reference level, and contrast the true and false displays with it.

We first implemented the maximal model with all relevant variables. We include both the main effects of *Display* and *Structure* and their interactions by including *Display* Structure* in the mean function. In the random structure, we include both random intercepts and random slopes for all grouping variables. We compare the model with simpler models by performing likelihood ratio tests using the anova function in R.

3.3. Results



When conjunctions are used, participants primarily gave the intermediate response when the

display is non-uniform, i.e., two of the cats were found, two of the cats were not found. A similar amount of participants gave the completely false response when the display is non-uniform. When the display shows that none of the four cats were found, the participants uniformly choose the completely false response. Similarly, when the display shows that all four cats were found, the participants uniformly choose the completely true response.

For sentence with *all*, participants almost uniformly gave the completely false response when the display is non-uniform or all the cats are uniformly not found. When the display shows that none of the four cats were found, the participants uniformly choose the completely true response.

We fitted a mixed-effects cumulative link model to responses to critical items with Structure and Display as fixed effects, and random by-question slopes for Structure. For all comparisons, the p value is less than 0.001, and α is set to 0.05, indicating that we have strong evidence to reject the null hypothesis that the model does not have significant differences from simpler models. We thus use the following model as our final model.

(18) $\mathsf{Response} \sim 1 + \mathsf{Display} * \mathsf{Structure} + (1|\mathsf{Subject}) + +(1 + \mathsf{Structure}|\mathsf{Question})$

A summary of the coefficients is given below.

	Estimate	Std. Error	z value	Pr($>|z|$)
Structure 1	10.134	1.547	6.550	$5.74e^{-11}$
Display 0	-8.097	1.230	-6.581	$4.67e^{-11}$
Display 4	32.135	5.524	5.817	$6.00e^{-09}$

From the table, we can conclude that we have strong evidence for believing that the structural difference and display difference, especially between 4/4 and 2/4, are important in explaining the response variable.Comparison of the model between the maximal model and those without each of the fixed effects revealed a significant effect of Structure ($p < .001$), a significant effect of Display ($p < .001$).

3.4. Discussions

In Experiment 1, we compare positive sentences with nominal conjunctions and *all*-NPs. The result confirms previous predictions on *all*-NPs. Gaps were not observed for *all*-NPs. Under non-uniform contexts, participants uniformly choose *Completely false*. However, sentences with nominal conjunctions get gappy judgement under non-uniform contexts. More than half of the participants choose *Neither completely true nor completely false*, similar to what Križ and Chemla (2015) observed for definite plurals. The result supports Schwarzschild's proposal and challenges Magri's proposal.

4. Experiment 2

In this experiment, we manipulate monotonicity. We ask the question whether the homogeneity of definite conjunction is symmetric or non-symmetric in positive and negative polarity contexts.

4.1. Methods

4.1.1. Participants

42 participants were recruited through Prolific. Participants were prescreened for their native language (English) and location (USA). Participants were paid 1.5 USD for the 8-minute study, for a pay rate of 16 USD/hour.

4.1.2. Procedures

The procedure was the same as Experiment 1. Experiment 2 also involved a ternary judgment task and was implemented using the Qualtrics platform. Participants were directed from Prolific to the experiment on the Qualtrics site.

In the experiments, the participants received a background story about a toy finding game. In each case, a child was assigned to a room with four hidden waving lucky cats. The child was encouraged to find as many cats as possible. Each cat has a unique name. The result of the toy finding game was presented, with a sentence describing the result provided under the result. The participants then had to decide upon seeing the result how right the sentence had been. Following the method in ternary judgment tasks, we provided participants with three response options: *Completely false*, *Neither completely false nor completely true*, *Completely true*. Participants indicated their response by clicking the button that had the desired response.

4.1.3. Materials

As in Experiment 1, each of the trails involved a display containing four objects. We used a green circle around the object to indicate that the cat was found by the child, and a red circle with a line through it to indicate that the cat was not found by the child. The cat icons were counterbalanced for their positions and frequencies.

The critical sentences are as follows.

(19) Target sentences
[Name] found [Cat name 1], [Cat name 2], [Cat name 3] and [Cat name 4].
Example: Elizabeth found Max, Coco, Luna and Tom.

(20) Control sentences
[Name] didn't find [Cat name 1], [Cat name 2], [Cat name 3] and [Cat name 4].
Example: Nina didn't find Max, Coco, Luna and Tom.

The critical sentences are provided under three types of display. In the first kind of display, all four cats were found. In the second kind of display, two out of four cats were found, the remaining two were not found. In the third kind of display, all four cats were not found.

Finally, alongside the critical sentences, we also include some filler sentences. These involved cases where the objects were mentioned using numerals, or quantifiers like *many* and *a few*. Unlike Experiment 1, where we only involved assertive sentences, for Experiment 2 we also included their negative counterparts.

(21) Examples of Filler sentences: Upward Entailing
a. Nina found more than two cats.
b. Roberts found less than two cats.

c. Mari found many cats.
d. Kevin found a few cats.

(22) Examples of Filler sentences: Downward Entailing
a. Nina didn't find more than two cats.
b. Roberts didn't find less than two cats.
c. Mari didn't find many cats.
d. Kevin didn't find a few cats.

The negative filler sentences shown in (22), especially those involving *many* and *few*, do not have clearly true or false values. Within the pool of filler sentences, we select an equal amount of completely true, completely false, and neither true nor false filler sentences.

In total, participants received two training items, followed by a semi-randomized sequence of 15 targets (5 under each display), 15 controls (5 under each display) and 30 filler sentences, for a total of 60 experimental trials.

4.2. Data analysis

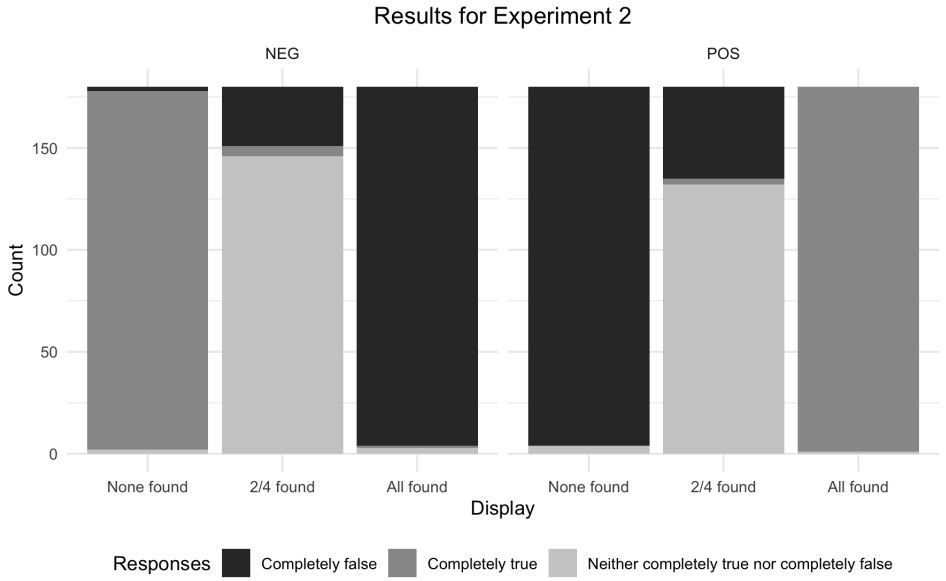
The factors we involve in the modeling include explanatory variables (display, monotonicity) and the response variable, which has three levels, i.e., *Completely true*, *Neither completely true nor completely false*, *Completely false*. The design of the experiment is largely the same as that of Experiment 1. Thus, again we fit the data to the Cumulative Link Mixed Model using the clmm function in the ordinal package in R. (Christensen, 2019)

In modeling, we treat *Response* as the response variable, *Display* and *Monotonicity* as the explanatory variables. We used contrast codings. For *Display*, there are three levels, i.e., 0/4 cats are found, 2/4 cats are found, 4/4 cats are found. We treat the non-uniform display, 2/4 as the reference level, and contrast the true and false displays with it.

We first implemented the full model with all relevant variables. We include both the main effects of *Display* and *Structure* and their interactions by including *Display* Structure* in the mean function. In the random structure, we include both random intercepts and random slopes for all grouping variables. We compare the model with simpler models and perform likelihood-ration tests using the anova function in R.

4.3. Results

The results of the positive sentences with conjunctions are as shown in the plot below. The participants uniformly gave the completely true response when positive sentences with conjunctions are provided in a context where all four cats were found. Participants uniformly gave the completely false response when positive sentences with conjunctions are provided in a context where all four cats were not found. When the display shows two of the four cats were found, two were not found, participants dominantly gave the neutral response, with a small portion choosing completely false.



For negative sentences, the participants uniformly gave the completely true response when negative sentences with conjunctions are provided in a context where all four cats were not found. Participants uniformly gave the completely false response when negative sentences with conjunctions are provided in a context where all four cats were found. When the display shows two of the four cats were found, two were not found, and participants dominantly gave the neutral response, with a small portion choosing completely false.

We fitted a mixed-effects cumulative link model to responses to critical items with Structure, Display and their interaction as fixed effects, and random by-participant slopes for Display and Structure, by-question for Structure. The result did not reveal a significant effect of Structure.

For all comparisons, the p-value is less than 0.001, α is set to 0.05, indicating that we have strong evidence to reject the null hypothesis that the model has no significant differences from simpler models. We thus use the maximal model as our final model.

(23) $\mathsf{Response} \sim \mathsf{Structure} * \mathsf{Display} + (1 + \mathsf{Structure} + \mathsf{Display}|\mathsf{Subject}) + (1 + \mathsf{Structure}|\mathsf{Question})$

A summary of the coefficients is given below.

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Structure 1	-0.7101	0.5371	-1.322	0.186
Display 0	9.0046	0.8320	10.823	$< 2e^{-16}$
Display 4	-7.5011	0.7934	-9.454	$< 2e^{-16}$
Structure 1:Display 0	-17.6342	1.4011	-12.586	$< 2e^{-16}$
Structure 1:Display 4	20.9582	2.0070	10.443	$< 2e^{-16}$

From the table, we can conclude that we have strong evidence to believe that the display difference is important in explaining the response variable. We lack evidence for believing that structure influences people's responses.

4.4. Discussions

In Experiment 2, we compare positive and negative sentences with nominal conjunctions. For both monotonicity, participants of similar percentages choose *Neither completely true nor completely false* under the non- uniform scenarios. The result shows that participants judge positive and negative sentences under non-uniform contexts symmetrically. The result supports Schwarzschild's proposal and challenges Magri's account of interference.

5. Discussions

5.1. Potential confounds

Before explaining the experiment design, we will mention two factors that we want to avoid influencing the results.

5.1.1. Non-maximality

It has long been noticed that under certain contexts, definite plurals are interpreted non-maximally. (Lasersohn, 1999) Brisson (1998) summarizes that the availability of non-maximality can be affected by a number of factors, including the lexical meaning and the size of the plurality.

(24) The boys arrived.

For example, for example, (24), in a scenario where one boy has not arrived, (24) is judged as more felicitous when there are many boys compared to when there are only three boys. There are different proposals on the nature of non-maximality, and debates on whether non-maximality is associated with homogeneity. (Lasersohn, 1999; Brisson, 1998; Bar-Lev, 2021; Križ and Spector, 2021; Sbardolini, 2023) We remain neutral about these questions, but we will try to control non-maximality so that it does not influence our experiment. We will do so by restricting the cardinality of all plurals in the experiment to 3. In addition, we will be using the same predicate.

5.1.2. Stressed *and*

Based on corpus data and eliciting judgment, Szabolcsi and Haddican (2004) observe that, in addition to the *neither* interpretation, English *and* has a marked *not both* reading. The availability of the reading depends mainly on the stress on the connective tissue, as shown in the following.

(25) a. Kate didn't found Max and$_F$ Tom.
b. Kate didn't found Max and Tom.

In (25a), *and* bears stress. The most readily available reading of (25a) is that Kate did not find both Max and Tom. In (25b), *and* does not bear stress. The reading is that Kate found neither Max nor Tom.

Another factor that may influence the interpretation of *and* is the semantic relatedness of the conjuncts. When the two conjuncts are not stereotypically associated with each other, a stressed *and* is strongly preferred over a non-stressed one.

(26) a. ?? Kate didn't take hockey and algebra.
b. Kate didn't take hockey or algebra.

c. Kate didn't take hockey or$_F$ algebra.

For instance, in the examples above, hockey and algebra are two things which are not commonly considered as related. According to Szabolcsi and Haddican (2004), when informants are given (26a), the majority judged that unless *and* is stressed as shown in (26c), the sentence is barely acceptable. (26c) interprets as Kate did not take both hockey and algebra. To achieve a *neither* reading, as intended in (26a), people will use (26b).

Our study is only interested in the default unstressed *and*. We will use the following three methods to prevent the interference of a stressed *and*.

We first make sure that the conjuncts are related. In addition, we put the target sentences into question-answer pairs as below.

(27) A: What did Kate find?
B: Kate found Max and Tom.
B': Kate found Max AND Tom.

The logic goes as follows. Rooth (1996) notices that the position of focus in an answer correlates with the questioned position in *wh*-questions. The questions determine sets of possible answers. The set of possible answers must match the alternatives of the answers. In (27), if we answer B' instead of B, the alternatives of the answer do not match the question, thus resulting in unnaturalness. Further discussion is available in the author's work in progress. To check interference with stressed *and*, we include sentences of the following structures and test the interpretation of the sentences of the people.

(28) Kate didn't find Max and Tom.

Individuals who interpret *and* with stress derive a "not both" reading of the sentence and judge the sentences as *Completely true* in non-uniform scenarios. As the results demonstrate, very few participants behaved in this way.

5.2. Križ and Chemla (2015)

The experimental study reduplicates the result presented in Križ and Chemla (2015). It should be noted that in Križ and Chemla (2015), although gappiness is presented in both positive and negative sentences with definite plurals, the percentages of people who give a neutral response are not the same (55% vs. 70 %). This contrast in percentage is not as obvious in our experiment. Although beyond the scope of this paper, in ongoing work the author is conducting a detailed comparative analysis of definite plurals and nominal conjunctions. Furthermore, Haslinger (2024) provides a thorough analysis of this puzzle and advances a proposed solution.

5.3. Homogeneity and non-maximality

Bar-Lev (2021) proposes that gappiness is a side effect of non-maximality. The proposal is supported by the gappiness asymmetry Križ and Chemla (2015) observed for definite plurals. Since nonmaximal readings are marginal under negation, so are the gaps in this environment. Križ and Spector (2021) proposes that the gapiness shows homogeneity (as true value gaps). No asymmetry for nonmaximality is predicted.

In our experiments, we show that nominal conjunctions show gappiness without allowing for

nonmaximality. There is thus a potential disassociation between the two phenomena.

Besides, we observe symmetry of gappiness between positive and negative contexts. We further challenge the association between nonmaximality and gappiness. The finding calls for more analyses on the relation between homogeneity and non-maximality. A comprehensive review and proposal can be found in Haslinger (2024).

5.4. More on the implicature account

Another influential implicature theory of homogeneity is Bar-Lev (2021). The implicature approach of the homogeneity of definite plurals assumes that the literal meaning of plural definites is weak. Thus, the strong reading under negation is part of the literal meaning. The maximal reading in the positive sentences is the result of implicatures, each subdomain alternaive is asserted by innocent inclusion. Note that Bar-Lev does not have predictions about conjunctions. The following discussion is an extension of Bar-Lev, assuming that conjunctions denote plural individuals like definite plurals. We will briefly explain how the theory works with the following example.

(29) a. Kate found the cats.
b. Kate didn't find the cats.

The literal meaning of (29a) is that Kate found one or more of the cats. In a scenario where Kate has two cats, *Max* and *Tom*, the literal meaning of (29a) can be represented below.

(30) $[\![(29a)]\!] = found(m)(k) \vee found(t)(k)$

The sentence is exhaustified, the implicature is as shown below.

(31) $[\![Exh^{IE+II}[\text{Kate found the cats}]]\!] \rightsquigarrow found(m)(k) \wedge found(t)(k)$

Exhaustification does not commonly take place under negation; thus the interpretation of (29b) is the same as its literal meaning as below.

(32) $[\![(29b)]\!] = \neg[found(m)(k) \vee found(t)(k)] = \neg found(m)(k) \wedge \neg found(t)(k)$

This is the interpretation of (29b). We summarize their predictions for conjunctions in Table 5.

(33) a. Kate found Max and Tom.
b. Kate didn't find Max and Tom.

Bar-Lev (2021) with $\oplus$	*found(m and t)*	$\neg$*found(m and t)*
$found(m)(k), found(k)(t)$	True	False
$\neg found(m)(k), found(k)(t)$	False (Implicature violation)	False
$found(m)(k), \neg found(k)(t)$	False (Implicature violation)	False
$\neg found(m)(k), \neg found(k)(t)$	False	True

Table 5: Bar-Lev (2021) predictions

The theory predicts that in a scenario where Kate found some but not all cats, (29a) has a true literal meaning but a false implicature.

Recently, Bassi et al. (2021) and Pinal et al. (2023) further develop mechanism of implicatures. Based on the theory, Guerrini and Wahbe (2023) gives a proposal on the homogeneity inference

of definite plurals. One can extend their proposals to the homogeneity of conjunctions. If one treats nominal conjunctions as $\oplus$, their proposal makes the predictions as in Table 6.

(34) a. Ann and Ben smiled.
b. Ann and Ben didn't smile.

(35) $pex([\![(34a)]\!]) = 1$ iff $\forall x : x \leq_{AT} a \oplus b \rightarrow x\ smiled$.
a. Assertion: $\exists x : x \leq_{AT} a \oplus b : x \in D \wedge x\ smiled$.
b. Presupposition: $\forall x : x \leq_{AT} a \oplus b \rightarrow x\ smiled \vee \neg\exists x : x \leq_{AT} a \oplus b \wedge x\ smiled$.

(36) $pex([\![(34b)]\!]) = 1$ iff $\neg\exists x \leq_{AT} a \oplus b : x\ smiled$.
a. Assertion: $\neg\exists x \leq_{AT} a \oplus b : x\ smiled$.
b. Presupposition: $\forall D' \subset D : D' \cap a \oplus b \neq \emptyset \rightarrow (\neg\exists x \leq_{AT} a \oplus b : x \in D' \wedge x\ smiled \vee$ $\forall x \leq_{AT} a \oplus b : x \in D' \rightarrow x\ smiled.)$

pex, $\oplus$	*found(m and t)*	$\neg$*found(m and t)*
$found(m)(k), found(k)(t)$	True	False
$\neg found(m)(k), found(k)(t)$	Presupposition failure	Presupposition failure
$found(m)(k), \neg found(k)(t)$	Presupposition failure	Presupposition failure
$\neg found(m)(k), \neg found(k)(t)$	False	True

Table 6: *pex* with $\oplus$ predictions

If one treats nominal conjunctions as $\wedge$, their proposal makes the predictions as in Table 7. As the tables show, the *pex* approach makes similar predictions as in Schwarzschild (1994). More discussions can be found in the author's work in progress.

(37) $Alt[\![(34a)]\!] = \{$a smiled, b smiled, a or b smiled$\}$

(38) $pex([\![(34a)]\!])$
a. asserts: a and b smiled.
b. presupposes: $\forall\alpha.\alpha \in R \wedge \alpha \in \{$a smiled, b smiled, a or b smiled$\} \rightarrow \alpha = 1 \vee$ $\forall\alpha.\alpha \in R \wedge \alpha \in \{$a smiled, b smiled, a or b smiled$\} \rightarrow \alpha = 0$

(39) $pex([\![(34a)]\!]) = 1$ iff *a and b smiled.*

(40) $pex([\![(34b)]\!]) = \neg pex([\![(34a)]\!]))$

(41) $\neg pex([\![(34a)]\!])$
a. asserts: $\neg$a and b smiled $= \neg$ a smiled $\vee \neg$b smiled
b. presupposes: $\forall\alpha.\alpha \in R \wedge \alpha \in \{$a smiled, b smiled, a or b smiled$\} \rightarrow \alpha = 1 \vee$ $\forall\alpha.\alpha \in R \wedge \alpha \in \{$a smiled, b smiled, a or b smiled$\} \rightarrow \alpha = 0$

(42) $\neg pex([\![(34a)]\!]) = 1$ iff *a didn't smile.* and *b didn't smile.*

5.5. Follow-up experiments

We also conducted follow-up experiments that compare definite numerals and conjunctions, both have been observed to be homogeneous but do not tolerate non-maximality. Although beyond the scope of this paper, in ongoing work the author discusses more on the results and issues raised.

pex, $\wedge$	*found(m and t)*	$\neg$*found(m and t)*
$found(m)(k), found(k)(t)$	True	False
$\neg found(m)(k), found(k)(t)$	Presupposition failure	Presupposition failure
$found(m)(k), \neg found(k)(t)$	Presupposition failure	Presupposition failure
$\neg found(m)(k), \neg found(k)(t)$	False	True

Table 7: *pex* with $\wedge$ predictions

6. Conclusion

In this paper, we report two experiments investigating how English-speaking adults interpret nominal conjunctions. The experiments reveal that participants exhibit gappy judgments for sentences with conjunctions, and these gaps are symmetric for positive and negative sentences in non-uniform contexts. These findings align with a trivalent approach to homogeneity (Schwarzschild, 1994; Križ, 2015; Križ and Spector, 2021) and challenge the implicature account proposed by Magri (2014). Although the implicature approach might be salvaged via the *pex* mechanism (Bassi et al., 2021; Pinal et al., 2023; Guerrini and Wahbe, 2023), further studies are required to assess its viability.

Appendix: Supplementary material and test items

The test items, collected data, and R data processing scripts are available at the following link: `https://github.com/ziaren/X_Conj`. The experiment was supported by departmental funds from Vincent Homer. Ethical review and approval were waived by the Human Research Protection Office of the University of Massachusetts (Protocol Number: 5362).

References

Abrusán, M. and K. Szendröi (2012). Experimenting with the king of france. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera (Eds.), *Logic, Language and Meaning*, Berlin, Heidelberg, pp. 102–111. Springer Berlin Heidelberg.

Augurzky, P., G.-U. Frankfurt, M. Bonnet, G.-A.-U. Göttingen, R. Breheny, A. Cremers, C. Ebert, G.-U. Frankfurt, C. Mayr, G.-A.-U. Göttingen, J. Romoli, H.-H.-U. Düsseldorf, M. Steinbach, G.-A.-U. Göttingen, and Y. Sudo (2023). Putting plural definites into context. In *Proceedings of Sinn und Bedeutung 27*.

Bar-Lev, M. E. (2021, October). An Implicature account of Homogeneity and Non-maximality. *Linguistics and Philosophy 44*(5), 1045–1097.

Bar-Lev, M. E. and D. Fox (2020, September). Free choice, simplification, and Innocent Inclusion. *Natural Language Semantics 28*(3), 175–223.

Bassi, I., G. Del Pinal, and U. Sauerland (2021, August). Presuppositional exhaustification. *Semantics and Pragmatics 14*(11), 1–48.

Beck, S. (2001). Reciprocals are definites. pp. 35.

Brisson, C. M. (1998). *Distributivity, Maximality, and Floating Quantifiers*. Ph. D. thesis, Rutgers University.

Christensen, R. H. B. (2019). *ordinal - Regression Models for Ordinal Data*.

Cooper, R. (1983). *Quantification and Syntactic Theory*. Springer.

Gajewski, J. R. (2005). Neg-Raising: Polarity and Presupposition.

Gazdar, G. (1980). A cross-categorial semantics for coordination. *Linguistics and Philosophy 3*(3), 407–409.

Guerrini, J. and J. Wahbe (2023). Homogeneity as pex.pdf.

Haslinger, N. (2024). *Pragmatic constraints on imprecision and homogeneity*. Ph. D. thesis, Universität Göttingen.

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

Keenan, E. L. and L. M. Faltz (1984). *Boolean Semantics for Natural Language*. Dordrecht, Netherland: Springer.

Križ, M. (2015). *Manuel Križ: Aspects of Homogeneity in the Semantics of Natural Language, c February 2015*. Ph. D. thesis, University of Vienna.

Križ, M. (2016, August). Homogeneity, Non-Maximality, and *all*. *Journal of Semantics 33*(3), 493–539.

Križ, M. and E. Chemla (2015, September). Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics 23*(3), 205–248.

Križ, M. and B. Spector (2021, October). Interpreting plural predication: homogeneity and non-maximality. *Linguistics and Philosophy 44*(5), 1131–1178.

Lasersohn, P. (1999). Pragmatic halos. *Language 75*(3), 522–551.

Löbner, S. (2000). Polarity in Natural Language: Predication, Quantification and Negation in Particular and Characterizing Sentences. *Linguistics and Philosophy 23*, 213–308.

Magri, G. (2014). *An Account for the Homogeneity Effect Triggered by Plural Definites and Conjunction Based on Double Strengthening*, pp. 99–145. London: Palgrave Macmillan UK.

Partee, B. and M. Rooth (1983). *Generalized Conjunction and Type Ambiguity*, pp. 361–383. Berlin, Boston: De Gruyter.

Pinal, G. D., I. Bassi, and U. Sauerland (2023). free choice and presuppositional exhaustification.

Renans, A., J. Romoli, M. M. Makri, L. Tieu, H. De Vries, R. Folli, and G. Tsoulas (2018, October). The abundance inference of pluralised mass nouns is an implicature: Evidence from Greek. *Glossa: a journal of general linguistics 3*(1).

Rooth, M. (1996). *The Handbook of Contemporary Semantic Theory*, Chapter Focus. Oxford: Blackwell.

Sbardolini, G. (2023). Homogeneity and Rejection.

Schwarzschild, R. (1994, October). Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics 2*(3), 201–248.

Schwarzschild, R. (1996). *Pluralities*.

Szabolcsi, A. and B. Haddican (2004, August). Conjunction Meets Negation: A Study in Cross-linguistic Variation. *Journal of Semantics 21*(3), 219–249.

Tieu, L., C. Bill, and J. Romoli (2017).

Tieu, L., M. Križ, and E. Chemla (2019, November). Children's Acquisition of Homogeneity in Plural Definite Descriptions. *Frontiers in Psychology 10*, 2329.

Tieu, L., J. Romoli, P. Zhou, and S. Crain (2016, May). Children's Knowledge of Free Choice Inferences and Scalar Implicatures. *Journal of Semantics 33*(2), 269–298.