An informativity-based account of negation complexity

Ming Xiang[a,*], Alex Kramer[b], Ann Nordmeyer[c]

[a] *Department of Linguistics, University of Chicago*

[b] *Department of Linguistics, University of Michigan*

[c] *Department of Psychology, Southern New Hampshire University*

Abstract

In sentence comprehension, negative sentences tend to elicit more processing cost than affirmative sentences. A growing body of work has shown that pragmatic context is an important factor that contributes to negation comprehension cost. The nature of this pragmatic effect, however, is yet to be determined. In four behavioral experiments, the current study assesses two possible pragmatic accounts: the expectation-based and the informativity-based accounts. Our findings suggest that informativity, instead of contextual expectation, is more directly responsible for negation comprehension. Contextual expectation only modulates negation comprehension cost if it facilitates the appropriate type of question under discussion.

Keywords: negation, QUD, pragmatics, informativity, contextual expectation

1. Introduction

In sentence comprehension, negative sentences tend to elicit more processing cost than affirmative sentences. Sentence verification tasks (Wason, 1959, 1961; Clark & Chase, 1972) have shown that negative sentences are generally more difficult for subjects to verify than their affirmative counterparts. Studies that have employed online comprehension techniques, such as ERPs (Fischler, Bloom, Childers, Roucos & Perry, 1983; Kounios & Holcomb, 1992), have also shown that negation does not seem to be evaluated soon enough to influence the N400 amplitude of an upcoming target word. Sentence pairs like *A robin is/is not a bird* (Fischler, Bloom, Childers, Roucos & Perry, 1983), despite obvious truth value differences,

*Corresponding author

Email address: mxiang@uchicago.edu (Ming Xiang)

did not produce N400 differences on the critical word *bird*, suggesting lexical semantic associations (e.g. *robin–bird*), instead of the truth value of the sentence, modulated the N400 amplitude in this case.

Several possible explanations exist to explain the processing cost of negation. One earlier hypothesis was that the comprehension of a negative proposition is decomposed into stages: the affirmative subject-predicate relation is processed first, and after that, the negative logical relation is processed in the second stage (Clark & Chase, 1972; Gough, 1965; Trabasso, Rollins & Shaughnessy, 1971), leading to extra cost, as well as delayed processing of negation.

There are, however, both theoretical and empirical challenges to a two-stage account of negation comprehension. Theoretically, a two-stage account runs counter to the standard view that language comprehension is highly incremental. Empirically speaking, a number of studies have shown that the processing of negation is not necessarily delayed. Nieuwland & Kuperberg (2008) argued that negation is only pragmatically licensed when it is used to reject what plausibly may have been true (i.e. *plausible denial*, Wason 1965). They found that, once the pragmatic condition on using negation was met, sentential negation was incrementally processed and can influence the N400 on the upcoming word. Making use of the grammatical dependency relation between negation and negative polarity items (e.g. the English word *ever*), an ERP study in Xiang et al. (2016) also demonstrated that negation can be processed quickly enough online to influence the N400 amplitude on an upcoming negative polarity item (e.g. *No hotels that the travel books included in their latest editions have ever received serious complaints*). Tian, Breheny & Ferguson (2010) and Tian, Ferguson & Breheny (2016) found that changing the linguistic input from a simple negative sentence (e.g. *Mike didn't iron his shirt*) to a negated cleft structure (e.g. *It was Mike who didn't iron his shirt*) helped participants to rapidly access the content of the negative assertion. For example, while hearing the simple negative sentence *Mike didn't iron his shirt*, during the post-verb time window, participants looked at both the visual object representing the positive state of affairs (*the ironed shirt*) and the visual object representing the negative state of affairs (*the crumpled shirt*). But with the negated cleft structure, participants' eye gaze showed no delay to the target picture representing the content of the negative assertion (*the crumpled shirt*). They therefore concluded the comprehension of negative sentences is not always delayed, instead, it depends on the implicit question under discussion in the discourse.

Thus, there is growing evidence that the processing costs of negation can be reduced in the appropriate context. The nature of this context effect, however, is still under debate. The current study aims to further our understanding of the pragmatic constraints on negation comprehension by comparing two different hypotheses regarding the source of the complexity of negation: the contextual expectation-based account and the informativity-based account.

The source of negation complexity: Contextual expectation of the negated property

A widely cited proposal from Wason (1965) is that the difficulty involved in denying a property is a function of how easily the positive property can be perceived. This proposal could be interpreted as suggesting that contextual expectation of the relevant positive property modulates the processing cost of negation. This explains why denying a property that is hardly expected to begin with, e.g. denying the proposition *A robin is a tree* to form a negative true proposition *A robin is not a tree*, will be costly. This interpretation is also in line with an independent line of proposals that suggested negative propositions presuppose their affirmative counterparts (Baldwin, 1928; Horn, 1989; Givón, 1979). As an example from Givón (1979), the utterance *My wife is not pregnant* is only pragmatically felicitous if there is already a shared/presupposed belief (or expectation) in the interlocutors' common ground that the speaker's wife could have been pregnant.

If contextual expectation *pragmatically licenses* negation, one may predict that the processing cost of negation is a function of how expected the relevant positive property is in a context. A recent study from Nordmeyer & Frank (2014) systematically manipulated contextual expectation and indeed observed that the reaction time to evaluate negative propositions was proportional to listeners' expectation of the to-be-negated property. This study will form the basis for the current investigation, and we introduce here the design features and the findings that are the most relevant for the current purpose. In N&F's study, after being exposed to a context scene, participants made a truth-value judgment on a target sentence. For instance, in Figure 1, participants first viewed a context scene with three boys, none of whom had apples in their hands. Then in the target picture-sentence scene, they saw a new boy with some apples and a sentence that said *Bob has no apples*. Participants then made a force-choice truth value judgment on the target sentence.

The experiment manipulated the polarity of the target sentence (affirmative or negative),

the truth condition of the target sentence (a true or false description of a given target picture), and most critically, the context scenes systematically varied the expectation of a relevant positive property. As shown in the example stimuli set presented in Figure 2, if the target sentence was *Bob has apples/no apples*, the context varied from 0 (out of 3) boys with apples to 3 boys with apples, and the target trial paired an affirmative or negative description with a picture in which a boy either had apples in his hand or had nothing. There were a total of 16 different conditions, combining different context scenarios and target pictures/sentences.

Figure 1: A sample trial from Nordmeyer & Frank (2014)

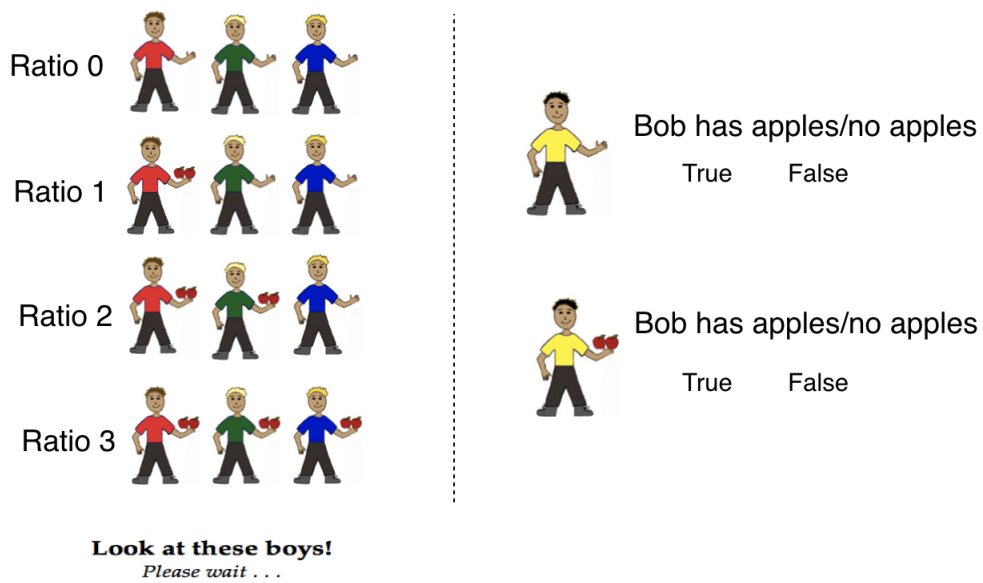

Figure 2: A full stimuli set from Nordmeyer & Frank (2014)

In Figure 2, the contextual expectation of the positive property "a boy has apples" was operationally defined as the ratio of the apple-possessing boys in the context scene. The critical finding from this study was that the time it took for participants to evaluate the

true negative sentences was negatively correlated with the ratio of the apple-possessing boys in the context: when the number of apple-possessing boys increased in the context, the time taken to judge the negative sentence *Bob has no apples* decreased. Such findings lend support to the hypothesis that the contextual expectation of the positive property plays an important role in explaining the comprehension difficulty associated with negation.

The source of negation complexity: (un)informativeness of negative utterances

Different from an expectation-based account, another account of negation complexity is grounded in the notion of informativity. Horn (1989), citing Leech (1981), suggested *the principle of negative uninformativeness*. This principle argues that negative propositions are in general far less informative than affirmative ones, and therefore everything being equal speakers will not use negative utterances. Crucially, though, negative propositions become informative when the corresponding affirmative propositions are expected. We agree with the general proposal from Horn. However, it wasn't clear from Horn's claims how and why the informativeness of the negative propositions should be associated with the expectation of the corresponding affirmative propositions. The current study aims to probe this intuition further and by doing so provides a more precise account of the difficulty associated with the comprehension of negation. To preview the upcoming discussion, while we think the informativity-based account is different from the expectation-based one, it is important to note that ultimately the notions of expectation and informativity can be made compatible with each other, but crucially under a very specific analysis about how to evaluate whether a negative utterance is informative or not in a discourse context. Many of the previous studies on negation have utilized both notions of expectation and informativity and used them almost interchangeably, but rarely have these studies clearly articulated the theoretical underpinning of these two constructs, and how they are different or related to each other. We will argue that the core notion responsible for negation complexity is informativity defined over the relevant discourse questions at issue; and that the apparent effect of contextual expectation of a positive property is ultimately subsidiary. More specifically, we will argue that contextual expectation only makes correct predictions if the question under discussion in the discourse is of the right type that renders a negative sentence an informative utterance.

Let's start by giving a formal definition of informativity. Following the recent work in Bayesian pragmatic reasoning (Frank & Goodman, 2012; Goodman & Frank, 2016), the informativeness of an utterance can be quantified as the negative surprisal (positive log

probability) of an intended message m given an utterance u (when the utterance is true), as shown in the equation below:

$$informativeness \propto log(P(m|[\![u]\!] = 1)) \tag{1}$$

The term *message* here is meant to be interchangable with the term *state of the world*. A listener, upon hearing an utterance u, probabilistically updates his beliefs about different possible states of the world. The set of the possible states that a listener entertains, however, is constrained by the communicative context. Specifically, we assume that discourse context is structured via explicit or implicit questions under discussion, or QUDs (Roberts, 1996; Farkas & Bruce, 2010; Ginzburg, 1996). These are simply salient discourse *issues* that engage the interlocutors in a communication. When a salient (implicit) question is raised in a context, there will be a set of possible answers to the question, with each answer representing a different (but possibly also overlapping) state of the world. When a speaker makes an utterance, a listener will use the utterance to update his beliefs about these states of the world. Consequently, information exchange happens and the discourse progresses forward. We note that the notion of QUD is used here as a convenient theoretical tool to more concisely capture what the discourse is about at any given moment, but it is not meant to claim that discourse participants (the speaker and the listener) necessarily construct actual explicit questions before processing an utterance.

To demonstrate why negative propositions are usually uninformative, let's consider a hypothetical situation in which the salient discourse QUD is *What does John have?*. We will call this type of question *Alternative Questions*, since the potential messages that could address this question form a set of alternative propositions in the form of *John has X*, such as {*John has apples, John has pears, John has oranges,* etc}. If a speaker knows that the actual answer to this question is John has apples, the best strategy is to utter the affirmative utterance *John has apples*. Since this utterance selects a single proposition from the space of possible answers and removes all other alternative messages, it therefore maximizes the probability that a listener will correctly receive the intended message and successfully update her discourse model. In contrast to this, a negative utterance such as *John doesn't have pears* or *John has no pears*, although compatible with the intended message *John has apples*, only removes one proposition (i.e. *John has pears*) from the set of all alternative propositions. The listener therefore can only obtain very limited information from this utterance to update her beliefs, and the probability that she will receive the intended message remains small. This

simple example suffices to demonstrate the more general point that negative propositions are often uninformative compared to affirmative propositions.

In addition, a rational speaker should make utterances that not only maximize informativity for the listener but also minimize utterance cost (Grice, 1975; Horn, 1989; Frank & Goodman, 2012). Everything else being equal, negative utterances are often more costly to produce due to the additional morpho-syntactic components associated with negation. If negative sentences are both uninformative and costly, it follows then speakers should prefer not to use negation, and comprehending these uninformative negative utterances will incur additional processing difficulty.

Under some very specific types of discourse QUDs, however, negative utterances do not necessarily lose the competition with their affirmative counterparts. Consider a QUD *Does John have apples?*. We will call this type of questions *Polar Questions*, since the most salient set of possible answers that can address this question includes two opposite propositions {*John has apples*, *John has no apples*}. An affirmative utterance will uniquely identify one message from this set, and a negative utterance will uniquely identify the other. Under a polar question, therefore, the affirmative and negative utterances are equally informative to deliver the respective target message.

Comparing the *Polar Question* with the earlier example of the *Alternative Question*, a negative utterance only stands the chance to be informative when addressing a *Polar* QUD. Since a Polar question would most naturally arise if the context makes a single positive property salient and there are questions about whether this property holds or not, this would explain the intuition that negative utterances are most felicitous when the contrastive positive property is expected or easily perceivable in the context. But under the informativity-based account, what really *licenses* negative utterances is not the expectation per se, but the general pragmatic principle that derives the informativeness of an utterance relative to a discourse QUD.

The expectation-based and the informativity-based accounts therefore make different predictions. Under the expectation-based account, the salience or expectation of a positive property directly modulates the processing costs of negation, and therefore one should predict an across-the-board facilitation effect of contextual expectation. Under the informativity-based account, however, contextual expectation is only helpful if the relevant discourse QUD is of the appropriate type (i.e. a polar QUD). The current study aims to distinguish these two related, but distinct, hypotheses.

Current study

The current study is built upon the study from Nordmeyer & Frank (2014). In the original study, increasing the contextual expectancy/salience of a positive property, e.g. increasing the number of boys who have apples, led to reduced processing difficulty when a negative statement such as *Bob has no apples* was encountered. The design of this study conflates expectation and informativity, thus lending support to both the expectation based account of negation complexity and the informativity based account. In particular, in the scene depicted in Figure 2, the visual context creates a very salient contrast between the two groups of boys: there are boys who have apples and boys who do not. The most salient *issue* in this context thus seems to be best captured by the Polar Question *Does Boy X have apples (or not)?*. The space of possible answers contains two opposite propositions: {*X has apples; X doesn't have apples*}. As discussed earlier, given such a QUD and the space of salient possible answers, a negative utterance such as *Bob has no apples* is as informative as an affirmative utterance *Bob has apples*, since both sentences uniquely identify a state of the world that can speak to the current discourse question.

Consider now what happens if the context changes such that every boy in the context has something, but some have apples and some have oranges. Such a context seems to put more focus on what the boys have, instead of whether they have one particular kind of fruit or not. In other words, the relevant discourse QUD seems to be more naturally construed as *What does X have?* than as *Does X have apples or not?*. Given that there are two possible kinds of fruit a boy could have in this context, the space of possible answers to this question would be the following, corresponding to different states of the world: {*X has apples; X has oranges; X has both; X has nothing*}. If the speaker utters the negative sentence *X has no apples*, the listener will update her beliefs and remove world states incompatible with the utterance. The negative sentence itself is ambiguous, however, since there will be two remaining states left, {*X has oranges; X has nothing*}, both compatible with the literal meaning of the negative utterance *X has no apples*. The negative utterance, therefore, is an uninformative utterance for the listener, since the speaker could have chosen an unambiguous and thus more informative utterance, such as *X has oranges* or *X has nothing*, to help the listener to uniquely identify the intended message.

We call the first type of context, the one used in Nordmeyer & Frank (2014), *Polar Context*, and the second type *Alternative Context*. In Experiment 1 below, we replicate the N&F findings with their original material and paradigm. In Experiment 2, we replace the *Po-*

lar Context in N&F with an *Alternative Context*. In both experiments, we maintained the original design feature that manipulated the expectancy/salience of the affirmative proposition. The expectation-based account of negation comprehension predicts that participants in both experiments will show reduced processing difficulty when there is increased expectation/salience for the positive property, whereas the informativity-based account predicts an effect of contextual expectation of the positive property only for Experiment 1 (the Polar Context), but not for Experiment 2 (the Alternative Context).

2. Experiment 1: replication of N&F 2014

Material, procedure and participants

The Experimental design follows the same setup as shown in Figure 2. For each trial, participants were first presented with a visual context display. The context display varied in the ratio of boys having a particular property to boys without that property. For example, in Figure 2, the number of boys that have apples increased from 0 to 3. The visual context was presented for 4500 ms before the target image and the target sentence were presented. There were two possible target images: either a character that has the relevant property (e.g. having apples) or does not (e.g. having nothing). The affirmative or negative statement paired with the target image was therefore either true or false. The participants were instructed to make a truth value judgment. The design was largely identical to the original study in N&F (2014), with the exception that the *Ratio* manipulation in the visual context was a within-subject factor in the current study, whereas it was a between-subject factor in the original study. All other factors in the current study were also within-subject factors. Each item set therefore contained 16 different variations (4 Ratios, affirmative/negative target sentence, true or false target statement given the target image). A total of 128 self-reported native English speakers, recruited from Amazon Mechanical Turk participated in our study. We used jsPsych (De Leeuw, 2015) to collect reaction time measures on MTurk. Thirty-two item sets (each with 16 conditions) were adapted from the original stimuli in N&F(2014). Each participant only viewed one condition from each item set, with a total of 32 trials. Since different trials were drawn from different item sets, during the experimental session the characters in the visual contexts had different objects. The experimental session took about 10-15 minutes to complete, and each participant received a payment of $1-$1.5. The current study was approved by the Institutional Review Board at the University of Chicago.

We also note that the two hypotheses introduced earlier, the expectation-based and the informativity-based accounts, make concrete predictions only for trials in which the target sentences correctly describe the target images (e.g. the TRUE trials). The informativity metric defined in (1) is only meaningful for utterances that are TRUE descriptions of a given state of the world. In particular, we predict that for negative sentences that are, semantically-speaking, *TRUE*, the processing difficulty incurred is a function of the appropriate pragmatic support (expectation or informativity under the current two hypotheses) from the context. We will use the response time participants took to make their judgments as an index of such processing difficulty. When a negative statement is FALSE, however, participants may adopt different decision criteria and response strategies, and our current hypotheses do not speak to those issues directly. In our report of the findings, we will therefore only analyze the response times and the response accuracy for the semantically TRUE trials.

Results

Before data analysis, we removed 6 participants who were beyond 60 years old. For the remaining 122 participants (mean age 34 years, 65 females), we also excluded trials that took more than 6000ms for participants to make their judgments, which excluded about 4% of the data. The RT data was also log-transformed before statistical analysis.

The results are presented in Figure 3. Both the accuracy and the RT results (for only the semantically TRUE trials) are also plotted in Figure 10 below in a summary plot that plots findings from all experiments together. Participants were highly accurate in their verification of the target sentences. For the TRUE trials, the mean accuracy on the true affirmative trials was 0.96 $\pm$ 0.01, and on the true negative trials was 0.93 $\pm$ 0.01. We conducted a logistic mixed effects model on the responses from the semantically TRUE trials. The variable *Ratio* was treated as a continuous variable, and the variable *Polarity* (affirmative or negative target sentence) was treatment-coded (negative was coded as 0, and affirmative was coded as 1). The model included Ratio, Polarity and their interaction as the fixed effects; and the random effects structure included the by-participant and by-item intercepts [1]. There was no difference in accuracy between affirmative and negative trials (p>0.3), nor did different ratios

[1]Model specification: model=glmer(Response~Ratio*Polarity+(1|Participant)+(1|Item), data, family=binomial). Models with more complex random effects structure did not converge. Unless otherwise noted, all models reported in this paper included the maximal random effects structure that successfully converged.

in the context have an effect on accuracy (p>0.3), and there was no interaction between the two variables (p>0.2).

For response time, we excluded trials where the participant made inaccurate responses. A linear mixed effects model on the log-transformed RT data[2] revealed a significant effect of Ratio (Est=-0.02± 0.008, t=-2.4, p<.05). Overall, for both affirmative and negative target sentences, the response time linearly decreased as the ratio moved from the lowest to the highest in the context. This fully replicated the critical finding in N&F (2014). There is also a significant effect of sentence polarity, such that negative sentences took longer than affirmative sentences across the board (Est=-0.11± 0.02, t=-4.5, p<.0001). There was no interaction between Ratio in context and the target sentence polarity ((Est=0.007± 0.01, t=0.6, p>.5).

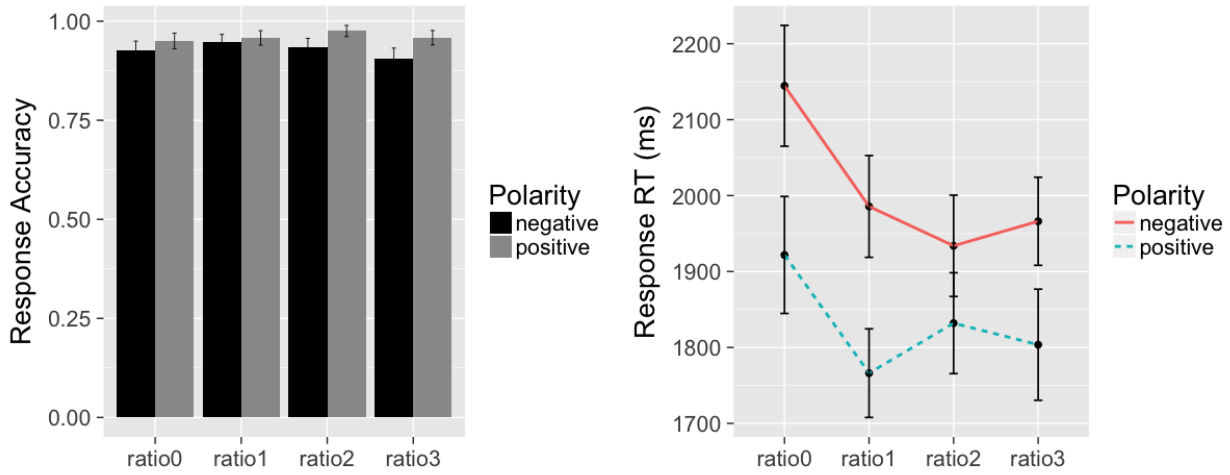

Figure 3: Response accuracy and response RTs for the semantically TRUE target sentences in Experiment 1. Error bars show standard error.

3. Experiment 2

Experiment 2 had a very similar design to Experiment 1. The only difference was that while Experiment 1 presented a *Polar Context* to the participants, Experiment 2 presented an *Alternative Context*. As shown in Figure 4, the visual context display made salient the distinction between two kinds of characters: ones that have apples and ones that have oranges.

[2]Model specification: model=lmer(lgRT~Polarity*Ratio+(1+Polarity*Ratio|Participant)+(1+Polarity*Ratio|Item), data)

As discussed earlier, in comparison to Experiment 1, the *Alternative Context* encourages participants to interpret the most relevant *issue* in the discourse as being about what object each character has. A reasonable QUD in this case could be *What does the boy have?*. Under such a QUD, a negative statement such as *Bob has no apples* is much less informative than an affirmative statement such as *Bob has apples*, because the former utterance leaves the listener with uncertainty as to which state of the world they should commit to. Even in situations where a character indeed has nothing and a negative utterance is required to describe the situation, as is the case for the target image depicted at the upper right corner of Figure 4, the most informative (and also less costly) negative utterance should be the unambiguous utterance *Bob has nothing*, rather than *Bob has no apples*. We therefore predict that, unlike Experiment 1, increasing the salience of the positive property (i.e. the ratio of *apple-possessing boys*) in the context will not have a strong effect on the response time to the TRUE negative trials in Experiment 2.

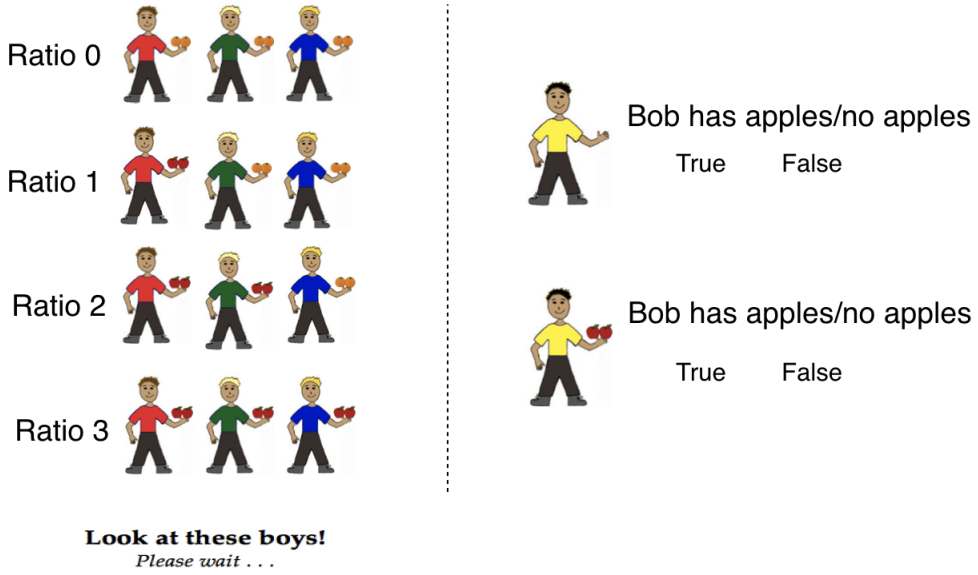

Figure 4: Sample stimuli for Experiment 2: Alternative Context

128 participants completed our experiment on Mechanical Turk. We excluded 1 participant whose self-reported dominant language was not English, and also excluded an additional 9 participants who were beyond 60 years old. For the remaining 118 participants (mean age 32 years old, 59 females), the data analysis procedure was identical to Experiment 1. We again only analyzed the response accuracy and response RTs to the semantically TRUE trials.

The results are presented in Figure 5. The mean accuracy for the true affirmative trials

was 0.97 ± 0.006, and 0.95 ± 0.01 for the true negative trials. The mixed effects logistic model on the responses included Ratio, Polarity and their interaction as the fixed effects, as well as the by-participant and by-item random intercepts[3]. The affirmative sentences were reliably more accurate than the negative trials (Est=0.8 ± 0.4, z=2.1, p<.05). The variable Ratio also had an effect (Est=0.33 ± 0.15, z=2.2, p<.05). There was no interaction between the polarity of the target sentence and the ratio factor from the context (Est=-0.06 ± 0.2, z=-0.2, p>.8).

We conducted a linear mixed-effects regression model on the log-transformed response time data[4]. There was no reliable effect of sentence polarity (Est=-0.04 ± 0.02, t=-1.6, p>.1). Different ratios in the context also had no reliable effect (Est=-0.015 ± 0.009, t=-1.7, p>.09). There was also no interaction between context ratio and target sentence polarity (Est=-0.002 ± 0.01, t=-0.1, p>.8).

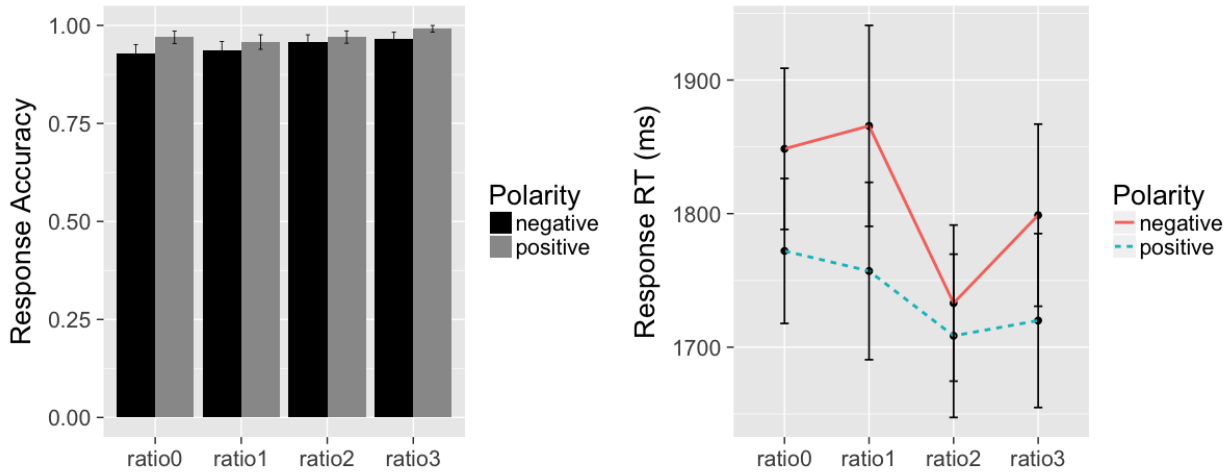

Figure 5: Response accuracy and response RTs for the semantically TRUE target sentences in Experiment 2. Error bars show standard error.

Summary of Experiment 1 and 2

Experiment 1 and 2 shared similar designs, but there was also a crucial difference between the two in terms of the information present in the context. The visual context in Experiment

[3]Model specification: model=lmer(Response~Polarity*Ratio+(1|Participant)+(1|Item), data, family=binomial)

[4]Model specification: model=lmer(lgRT~Polarity*Ratio+(1+Polarity*ratio|Participant)+(1+Polarity*Ratio|Item), data)

1 made a salient contrast between a property P (e.g. having apples) and the absence of P (e.g. not having apples); while the visual context in Experiment 2 focused on the contrast between a property P (e.g. having apples) and an alternative property Q (e.g. having oranges). Depending on which contrast was made salient in the visual context, a pragmatic listener should have made inferences about a different *Question under discussion (QUD)*, which represents the most relevant *issue* in the discourse context. Any new statement following the context scenario is only pragmatically felicitous if it properly addresses the implicit discourse QUD. We hypothesized that a negative statement is far less informative than an affirmative statement when addressing an Alternative QUD, but that a Polar QUD can be properly addressed by a negative statement.

The RT results showed that for a negative statement in the form of *X has no P*, processing difficulty decreased as the salience of the positive property *X has P* increased in the visual context, but crucially only when the context supported a Polar QUD (Experiment 1), not when the context supported an Alternative QUD (Experiment 2). This finding lends support to our hypothesis that it is the discourse QUD, not the expectancy of the positive property per se, that modulates the comprehension difficulty of negation. In particular, it is the informativity of a negative sentence in response to a specific QUD that accounts for its comprehension cost. A negative sentence is much less informative than its affirmative counterpart in many contexts, but it becomes informative when addressing a Polar QUD.

We want to note that the Ratio 3 visual context yielded identical conditions for experiment 1 and experiment 2. The Ratio 3 context presented scenarios in which all characters have the relevant positive property (e.g. all the boys are apple-possessing boys). At the individual trial level, the trials under the Ratio 3 context were identical for both experiments. In the global experimental context, however, participants in Experiment 1 were exposed to visual contexts that made a distinction between a polar contrast, i.e. having a particular object vs. having nothing; whereas participants in Experiment 2 were exposed to visual contexts that made a distinction between two alternatives, i.e. having one kind of object vs. having another kind. Given the different findings in Experiment 1 and 2, it is reasonable to assume that the global experimental context had a dominant effect in shaping the relevant QUDs. But we also acknowledge that under the Ratio 3 context there may be more variability at the individual trial level as to what QUD was the most appropriate.

Experiment 1 and 2 manipulated the property of the global discourse context to change the informativeness of a negative sentence. There are other factors that can change the

informativeness of a negative utterance as well. For example, the TRUE target negative sentences in Experiment 1 and 2 describe a character that has nothing in his/her hand. But what if the character in the target image possesses a different object? Intuitively, if Bob has an orange, a negative sentence *Bob has no apples*, although still a true description of Bob, is not the most informative description of the target character. The affirmative sentence *Bob has an oranges* should be used instead.

In Experiments 3 and 4, we examine whether the informativeness of a negative statement with respect to the local message influences its comprehension complexity. In these experiments, we kept the visual context displays the same as Experiments 1 and 2: Experiment 3 has the same Polar QUD context as Experiment 1, and Experiment 4 has the same Alternative QUD context as Experiment 2. However, we changed the target images that the negative sentences were a true description of, rendering the TRUE negative target sentences uninformative with respect to the target images. We make two predictions for experiment 3 and 4. First, we expect to replicate the effect of the global discourse QUD context. With a Polar QUD context, Experiment 3 should pattern like Experiment 1, under which increased salience of a positive property facilitates the processing of a negative utterance. On the other hand, no facilitation effect is expected under an Alternative QUD context in Experiment 4, parallel to Experiment 2. Second, we may expect an overall increase of processing difficulty on the target negative sentences in Experiment 3 and 4 compared to Experiment 1 and 2, since the target negative sentences are now much less informative with respect to the local target images.

4. Experiment 3

The design of Experiment 3 is demonstrated in Figure 6. The visual context display, like Experiment 1, sets up a polar contrast between a property P and the absence of the property P (i.e. having apples and not having apples in Figure 6). In the to-be-described target display, a character has either apples or oranges, and each target image is paired with either an affirmative or a negative statement.

The experimental procedure was identical to the previous experiments. 127 participants completed our experiment on Mechanical Turk. After excluding 5 participants were beyond 60 years old, data from 122 participants were entered into data analysis (mean age 36 years old, 70 females). The data analysis procedure was identical to Experiments 1 and 2.

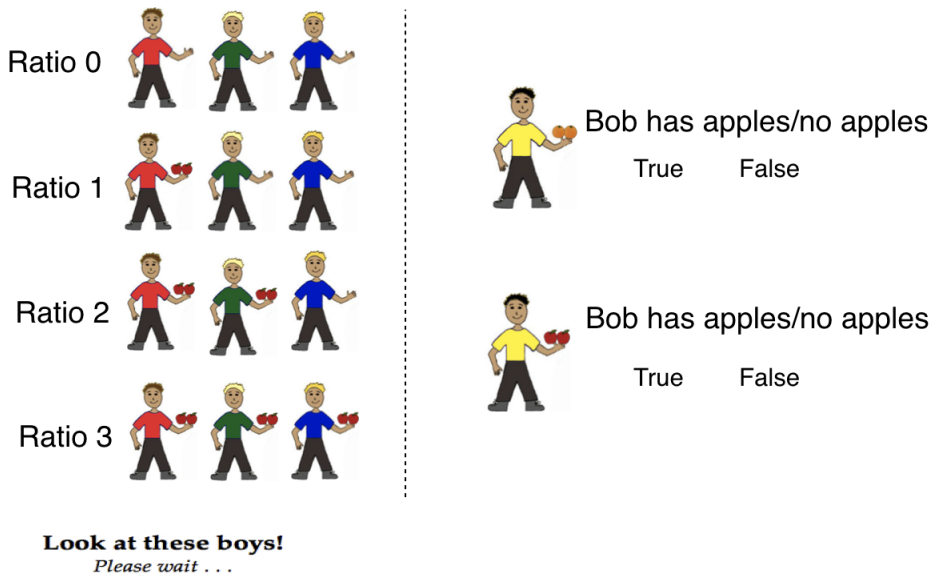

Figure 6: Sample stimuli for Experiment 3

The results are presented in Figure 7. The mean accuracy on the true affirmative trials was 0.96 ± 0.009, and 0.84 ± 0.02 for the true negative trials. The logistic mixed effects model[5] found that the affirmative sentences were more accurate than the negative sentences (Est=1.7 ± 0.4, z=5.2, p<.00001). The ratio variable did not have an effect (Est=0.06 ± 0.09, z=0.7, p>.5). There was no interaction between the polarity of the target sentence and the ratio factor in the visual context (Est=0.1 ± 0.18, z=0.5, p>.6).

For the response time, the linear mixed effects model on the log-transformed RTs[6] found a significant effect of sentence polarity, such that affirmative sentences were faster to verify than negative sentences across the board (Est=-0.29 ± 0.02, t=-11, p<.00001). There was also a significant effect of Ratio (Est=-0.02 ± 0.008, t=-2.1, p<.05), showing that for both affirmative and negative target sentences, the response time linearly decreased as the number of apple-possessing-boys increased from the lowest to the highest in the context. There was no interaction between Ratio and target sentence polarity (Est=-0.001 ± 0.01, t=-0.09, p>.9).

Experiment 3 replicated the Ratio effect in Experiment 1: under a Polar context, the process-

[5]Model specification: model=glmer(response~Ratio*Polarity+(1|Participant)+(1|Item), data, family=binomial)

[6]Model specification: model=lmer(lgRT~Polarity*Ratio+(1+Polarity*ratio|Participant)+(1+Polarity+Ratio|Item), data)

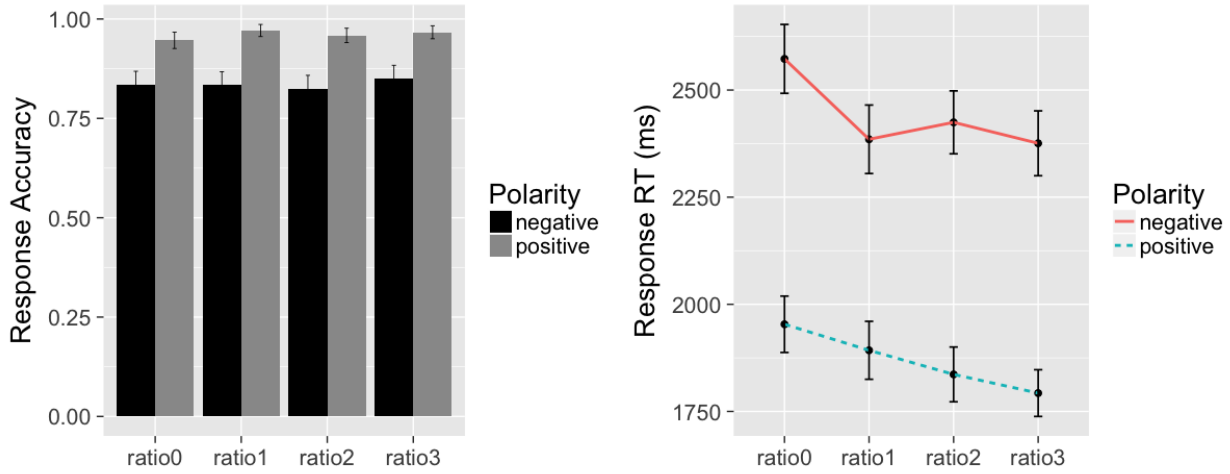

Figure 7: Response accuracy and response RTs for the semantically TRUE target sentences in Experiment 3. Error bars show standard error.

ing difficulty of negation decreased as the salience of the relevant positive property increased in the context. On the other hand, response times for the TRUE negative trials were much higher in Experiment 3 than in Experiment 1 (see the comparison in the summary plot in Figure 10), suggesting an overall increase in comprehension difficulty. We will discuss the cross-experiment comparison in more detail below.

5. Experiment 4

The design of Experiment 4 is demonstrated in Figure 8. The target images and sentences were identical to those of Experiment 3. However, the visual context scenario contains the Alternative Context from Experiment 2, which makes two alternative properties salient (i.e. *having apples* and *having oranges* in Figure 4).

127 participants completed our experiment on Mechanical Turk. After excluded participants whose self-reported dominant language was not English, and participants who were beyond 60 years old, a total of 123 participants were included in the data analysis (mean age 34 years old, 51 females). The data analysis procedure was identical to the previous experiments.

The results are presented in Figure 9. The mean accuracy for the true affirmative trials was

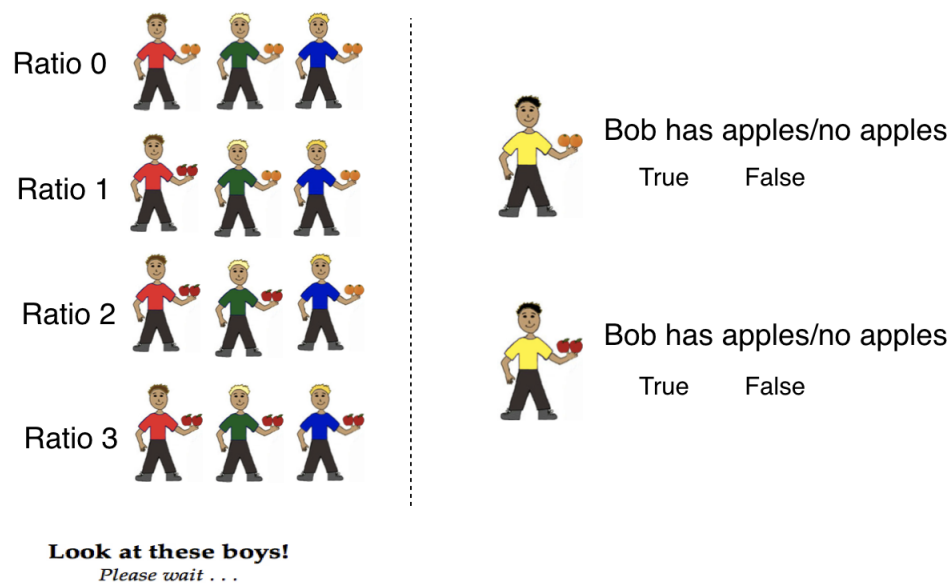

Figure 8: Sample stimuli for Experiment 4

0.95 ± 0.02, and 0.8 ± 0.02 for the true negative trials. The logistic mixed effects model[7] found that the affirmative trials were significantly more accurate than the negative trials (Est=1.7 ± 0.3, z=5.1, p<.00001). The ratio factor in the context did not have an effect (Est=0.04 ± 0.09, z=0.5, p>.6). There was no interaction between the polarity of the target sentence and the ratio factor from the context (Est=0.3 ± 0.2, z=1.5, p>.1).

The linear mixed effects model on the response RTs[8] found a significant effect of sentence polarity, such that negative sentences took longer than affirmative sentences across the board (Est=-0.2 ± 0.02, t=-8.5, p<.00001). Different ratios in the context had no reliable effect (Est=0.005 ± 0.008, t=0.6, p>.5). There was also no interaction between context ratio and target sentence polarity (Est=-0.02 ± 0.01, t=-1.6, p>.1).

Experiment 4 therefore replicated the main finding in Experiment 2, namely, that when the context scenario is an Alternative Context, the salience of a positive property in the context does not significantly affect the comprehension complexity of a counterpart negative property. However, we again also observed that response times for the TRUE negative trials were much higher in Experiment 4 than in Experiment 2 (again see the comparison in the

[7]Model specification: model=glmer(response~Ratio*Polarity+(1|Participant)+(1|Item), data, family=binomial)

[8]Model specification: model=lmer(lgRT~Polarity*Ratio+(1+Polarity*ratio|Participant)+(1+Polarity+ratio|Item), data)

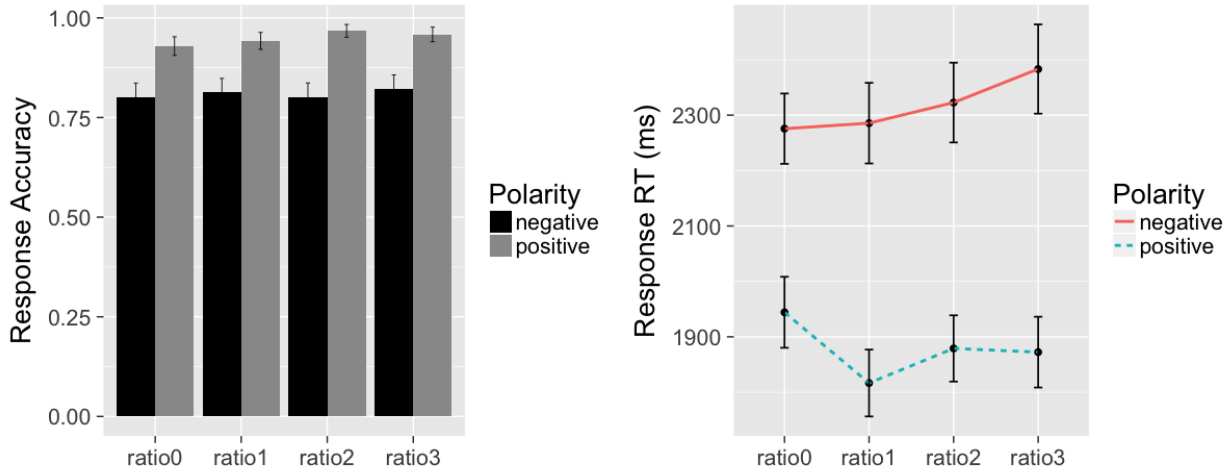

Figure 9: Response accuracy and response RTs for the semantically TRUE target sentences in Experiment 4. Error bars show standard error.

summary plot in Figure 10).

Summary of all four experiments

The results for all 4 experiments are plotted in Figure 10. These four experiments are different in terms of two design features. One feature is the visual context display: the visual context scenario either presented a Polar Context, which encouraged a Polar QUD (Experiment 1 and 3); or an Alternative Context, which encouraged an Alternative QUD (Experiment 2 and 4). The other feature is the target display: we labeled target displays in Experiment 1 and 2 as *Non-existence* in Figure 10, since the TRUE negative sentences corresponded to target images in which the character possessed no objects; conversely, we labeled target displays in Experiment 3 and 4 as *Alternative object*, since the TRUE negative sentences corresponded to target images in which the character possessed an alternative object. We conducted an analysis for response accuracy and response time on all data combined together. For response accuracy, the logistic regression model included Ratio, Polarity (affirmative vs. negative sentence), Context QUD (Polar QUD vs. Alternative QUD context) and Target image type (Non-existence vs. Alternative object) and their interactions as predictors. All predictors were sum-coded except that the variable Ratio is coded as a continuous variable. There was a significant interaction between sentence polarity and target image type (Est=-0.2 $\pm$ 0.08, t=-2.5, p<.01), driven by the fact that the accuracy on negative sentences was much lower than the affirmative sentences in Experiment 3 and 4 (with Alternative Object target image

type), compared to Experiment 1 and 2 (with Non-existence target image type).

For the analysis on response time, the mixed-effects regression model also included Ratio, Polarity, Context QUD, Target image type and their interactions as the fixed effects predictors, and it also included by-participant and by-item random intercepts, as well as the by-participant random slope with Ratio. There was a significant effect of Polarity (Est=0.08 $\pm$ 0.005, t=14, p<.00001). By and large, negative sentences took participants longer time to respond than affirmative sentences. But this negative sentence disadvantage was further modulated by two factors. First, there was a Polarity x Target Image interaction (Est=0.04 $\pm$ 0.005, t=7.8, p<.0001), driven by the fact that negative sentences took much longer to respond than their affirmative counterparts in Experiment 3 and 4 (Alternative Object target image) than in Experiment 1 and 2 (Non-existence target image). Second, there was also a Polarity x Context QUD interaction (Est=-0.02 $\pm$ 0.005, t=-3.5, p<.001). We did an analysis on affirmative and negative sentences separately to further assess the effect of context QUDs on different sentence types. For the affirmative sentences, there was an effect of Ratio (Est=-0.02 $\pm$ 0.004, t=-4.2, p<.0001), suggesting that the RT taken to respond to an affirmative sentence in general decreased as the contextual expectation of a positive property increased. But there was no effect of context QUD type (Est=-0.008 $\pm$ 0.01, t=-0.5, p>.5), nor an interaction between Ratio and Context QUD type (Est=0.0001 $\pm$ 0.004, t=0.04, p>.9). This suggests that for the affirmative sentences, although the salience of a positive property in the context has a general effect, the effect is not modulated by the QUD type signaled by the context. On the other hand, for the negative sentences, in addition to an effect of Ratio (Est=-0.01 $\pm$ 0.004, t=-3.1, p<.01) and Target image type (Est=0.1 $\pm$ 0.01, t=7.9, p<.0001), there are crucially also an effect of Context QUD type (Est=-0.05 $\pm$ 0.01, t=-3.4, p<.001) and a marginal interaction between Ratio and Context QUD type (Est=0.007 $\pm$ 0.004, t=1.8, p<.08). This confirms the conclusions we drew earlier based on the results from each individual experiment: the salience of a positive property facilitates the processing of a negative sentence only when the context supports a Polar Question QUD (Experiment 1 and 3), but not when the context supports an Alternative Question QUD (Experiment 2 and 4).

Therefore, when all the data are considered together, there is clear evidence that the comprehension complexity of negation is modulated by how informative the negative sentence is. In particular, the current study reveals that the informativity of a negative sentence must be assessed both at the global level: how informative an utterance is as an answer to address an implicit, discourse salient QUD; and at the local level: how informative an utterance is

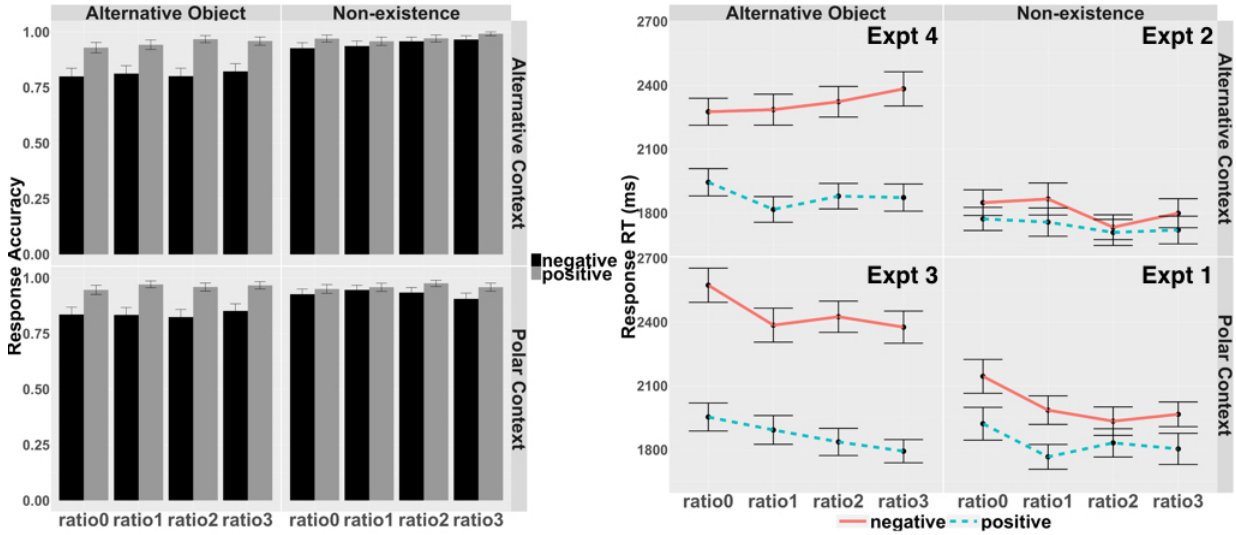

Figure 10: Summary of the results for Experiment 1-4. Accuracy is shown in the plots on the left, and reaction time is shown in the plots on the right. Bottom right: Experiment 1. Top right: Experiment 2. Bottom left: Experiment 3. Top left: Experiment 4. Error bars show standard error.

as a statement about a particular state of the world. We discuss these results in more detail below.

6. General Discussion

In four experiments, we showed that the processing difficulty of negation, as measured by the response time taken to evaluate a *true* negative utterance, is modulated by the informativeness of the negative utterance. At the global level, the informativeness of a negative utterance is determined by whether it constitutes an informative answer to an implicit discourse QUD. On the one hand, when the discourse context promotes a polar QUD (Experiments 1 and 3), a negative utterance can be added to the discourse as an informative utterance to address the QUD; reflecting this, we observed decreased processing difficulty as a function of increased salience of the corresponding positive property in the context. On the other hand, when the discourse context promotes an alternative QUD (Experiments 2 and Experiment 4), a negative utterance is not an informative utterance to address such QUDs, and the contextual salience of the corresponding positive property has no effect on the processing cost of negation. We also note that the current experimental manipulation at the global context level (Alternative vs. Polar context) is meant to manipulate the informativeness of the negative utterances only. We did not predict that comprehension of the affirmative sentences should

be affected by this particular type of QUD manipulation. It was indeed true, based on the analyses above, that the QUD type in the context did not have any reliable effect on the affirmative sentences. This is not to imply that the comprehension of affirmative sentences is in general insensitive to the informativeness condition. Instead, an affirmative sentence is an informative utterance to address either a Polar or an Alternative question we manipulated in the current study. Under both types of context, the degree of expectancy of a positive property was inversely related to the comprehension cost of the affirmative sentence, as reflected by an across-the-board effect of the variable *Ratio* for the affirmative trials.

At the local level, the informativeness of a negative utterance is determined by whether it is the most informative utterance or if there are other, more informative expressions that could describe the target state of affairs. We saw in Figure 10 that participants took much longer to respond to a negative sentence if the negative sentence was used to describe an *alternative object* target image, as opposed to a *non-existence* target image. This was presumably because with the *alternative object* conditions, the most informative description of the target visual display should be an affirmative statement. The observed effect was particularly associated with negative sentences. As we reported earlier, when all experiments were considered together, there was a significant interaction between sentence Polarity and Target image type in both the accuracy and the response RT measures, driven by the fact that the Target image type had a much stronger effect on the negative than on the affirmative sentences.

We take the current findings to suggest that informativeness, instead of contextual expectation per se, is a more fundamental notion in explaining the comprehension complexity associated with negation. This is not to say that an expectation-based explanation is necessarily wrong. Instead, what we observe is that when the notion of contextual expectation is constructed through the lens of what is relevant to the discourse or what the current question under discussion (QUD) is, informativity and expectation make very similar predictions. In the current study, when the discourse context makes a polar QUD salient, and hence makes an upcoming negative utterance potentially informative, we observe a negative correlation between the degree of contextual expectation of a positive property and the processing difficulty of the corresponding negative utterance. As we discussed earlier, a polar QUD is generally formed based on a positive property (e.g. Do the boys have apples?). It is thus possible that the more salient a positive property is in the discourse, the more likely participants are to form a polar QUD based on this positive property. In other words, in the right discourse, the increasing salience of a positive property in the context and the activation of a

polar QUD go hand-in-hand. However, when the dominant discourse QUD is an alternative question, a negative utterance is simply uninformative to address such QUDs, and hence we would observe little facilitation effect on negation even when there is strong expectation of a salient positive property in the context.

The current analysis is completely in line with the important insights from the proposals made by Tian, Breheny & Ferguson (2010) and Tian, Ferguson & Breheny (2016). The central question for Tian and colleagues was whether negation comprehension necessarily involves activating the events represented by the affirmative proposition. They argued that a simple negative sentence such as *NP has not P* is most likely associated with a QUD of the form *whether NP P*. To the extent that comprehenders are sensitive to such pragmatic constraints on negation, they would accommodate a polar whether-question for an isolated negative sentence. It is this QUD accommodation process that gives rise to the activation of affirmative propositions. But if a different QUD is provided (Tian and colleagues used cleft constructions to introduce different QUDs), one can eliminate the bias for affirmative propositions. The effect of QUD on negation processing has also been observed in language acquisition. For example, it has also been argued that when children evaluate the scope relation between a negation and a quantifier, they prefer an interpretation that forms an appropriate answer to a salient question under discussion (Gualmini, Hulsey, Hacquard & Fox, 2008; Musolino & Lidz, 2006).

Finally we note that although we found that the processing cost of negation is a function of its informativeness, enhancing informativeness in the current study did not entirely eliminate the difference between affirmative and negative sentences. As reported earlier, in all four experiments, by and large it consistently took participants longer to respond to negative sentences than to affirmative sentences, regardless of context and target display types. One possibility is that the contextual support for negative sentences was not sufficiently strong in the current study. There was no linguistic context for each experiment, and participants had to extract contextual information from highly simplified visual displays. We leave it open for future research about the strength of contextual support and what facilitates comprehenders to more effectively extract pragmatic information from context.

7. Conclusion

In four experiments, the current study aims to uncover the source of the apparent complexity of negation processing. Two hypotheses were compared to each other, one based on the

notion of informativity and the other on contextual expectation. We argue that informativeness is the primary source of negation complexity. Contextual expectation plays a role only when the discourse context renders a negative utterance informative.

References

Baldwin, J. M. (1928). Entries for negationand negative. *Dictionary of Philosophy and Psychology*, *2*, 146–49.

Clark, H., & Chase, W. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*, 1–12.

Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of semantics*, *27*, 81–118.

Fischler, I., Bloom, P., Childers, D., Roucos, S., & Perry, N. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, 400–409.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Ginzburg, J. (1996). Dynamics and the semantics of dialogue. *Seligman, Jerry, & Westerst ahl, Dag (eds), Logic, language and computation*, *1*.

Givón, T. (1979). *On understanding grammar* volume 379. Academic Press New York.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*, 818–829.

Gough, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of verbal learning and verbal behavior*, *4*, 107–111.

Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, *3*, 41–58.

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The question–answer requirement for scope assignment. *Natural language semantics*, *16*, 205.

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

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, *121*, 459.

Leech, G. N. (1981). Pragmatics and conversational rhetoric. *Possibilities and limitations of pragmatics*, (pp. 413–442).

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, *44*, 817–852.

Nieuwland, M., & Kuperberg, G. (2008). When the truth is not too hard to handle. *Psychological Science*, *19*, 1213.

Nordmeyer, A. E., & Frank, M. C. (2014). A pragmatic account of the processing of negative sentences. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics.ms. the ohio state university, .

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account.

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, *31*, 683–698.

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive psychology*, *2*, 239–289.

Wason, P. (1965). The contexts of plausible denial. *Journal of Verbal Learning & Verbal Behavior*, *4*, 7–11.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 92–107.

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*, 133–142.

Xiang, M., Grove, J., & Giannakidou, A. (2016). Semantic and pragmatic processes in the comprehension of negation: An event related potential study of negative polarity sensitivity. *Journal of Neurolinguistics*, *38*, 71–88.