Negation is only hard to process when it is not relevant or informative

Ann E. Nordmeyer

Department of Psychology, Southern New Hampshire University

Michael C. Frank

Department of Psychology, Stanford University

Author Note

Corresponding author: Ann E. Nordmeyer

School of Arts and Sciences

Southern New Hampshire University 2500 N. River Rd.

Hooksett, NH 03106

Phone: (603) 668-2211 x 2058

Email: a.nordmeyer@snhu.edu

Abstract

Negation is a fundamental element of language and logical systems, but processing negative sentences can be challenging. Early investigations suggested that this difficulty was due to the representational challenge of adding an additional logical element to a proposition, but in more recent work, supportive contexts mitigate the processing costs of negation, suggesting a pragmatic explanation. We make a strong test of this pragmatic hypothesis by directly comparing speakers and listeners. Speakers produce negative sentences more often when they are both relevant and informative. Listeners in turn are fastest to respond to sentences that they expect speakers to produce. Because negative sentences are only difficult in contexts when they are unlikely to be produced, the primary challenges in processing negation are likely due to general pragmatic principles that apply to all sentences, rather than representational factors specific to negation.

Keywords: Language, Psycholinguistics, Language Comprehension, Language Production, Pragmatics

Negation is only hard to process when it is not relevant or informative

Introduction

Language allows us to describe not only the world as we see it, but also the world as it is not. Nevertheless, for human language users, processing negation is often slow and effortful. Deciding the truth value of a sentence like "star isn't above plus" takes much longer than making the same decision about a positive sentence (H. Clark & Chase, 1972; Carpenter & Just, 1975; Just & Carpenter, 1971, 1976). And in language comprehension tasks, participants often appear to process the positive components of a sentence prior to negating them, suggesting again that negation is challenging (Kaup & Zwaan, 2003; Kaup, Ludtke, & Zwaan, 2006; Hasson & Glucksberg, 2006; Fischler, Bloom, Childers, Roucos, & Perry, 1983; Lüdtke, Friedrich, De Filippis, & Kaup, 2008; Ferguson & Sanford, 2008). Why do adults struggle to process negation despite producing negative sentences with ease?

One explanation is that not all negations are equally felicitous. For example, in many contexts it might be strange to say "my car isn't a minivan". And it would be even stranger to say "my car doesn't have an ejector seat." On Gricean and neo-Gricean accounts of language use in context, listeners expect speakers to produce truthful, informative, and relevant utterances (Grice, 1975; Horn, 1984; Levinson, 2000; Sperber & Wilson, 1986). The above utterances are both truthful, but vary in their informativeness and relevance.

Many formal theories of pragmatics focus specifically on the role of either relevance or informativeness (e.g. Sperber & Wilson, 1986; Frank & Goodman, 2012). In this paper, we examine the separate effect of these pragmatic factors. We define an informative utterance as one that conveys more information about the referent (i.e., makes the referent easier to identify in context; see Frank and Goodman (2012) for a formalization of informativeness, and H. H. Clark (1976) and Moxey (2006) for a discussion of informativeness as it applies to negative sentences). For example, the utterance "my car isn't a minivan" doesn't convey as much information as, say, "my car is a sedan", because "isn't a minivan" refers to a larger set of cars than "is a sedan". If we were in a parking lot full of minivans, however, the statement “my car isn’t a minivan” becomes much more informative, because it helps to uniquely identify the car in question *in that particular context*.

We define a relevant feature as one that is salient or expected based on the context. Although under this definition an informative utterance is *always* relevant (because any feature that helps you identify a referent will also be something noticeable or salient about that referent in context), a relevant utterance is not necessarily informative. For example, the sentence “my car isn’t a minivan” is relevant even though it is not usually informative, because the category that a car belongs to is a salient feature (i.e. many people will mention the type of car when asked to describe a particular car). The sentence “my car doesn’t have an ejector seat”, however, is neither relevant nor informative, because in most contexts having an ejector seat is not a feature that we would expect a car to have. If you are James Bond, however, and you live in a universe where ejector seats are sometimes a feature that cars have, then the utterance “my car doesn’t have an ejector seat” becomes very relevant (though still not particularly informative, because it is still the case that the *majority* of cars don’t have ejector seats). Thus, we take relevance and informativeness to be separable pragmatic factors, and both can change depending on the context in which an utterance is produced.

A key element of neo-Gricean theories of communication is that listeners *expect* speakers to produce utterances that are informative and relevant. These pragmatic expectations are present regardless of whether a sentence is affirmative or negative. When an utterance appears to violate these principles, listeners make inferences about the speaker’s intended meaning that go beyond the literal meaning of the utterance. Neo-Gricean theories have been used to explain the pragmatic inferences that can be drawn from a negative utterance (Moeschler, 1992, see Tian, Breheny, & Ferguson, 2010 for experimental support); for example, if I told you that my car isn’t a minivan, you might *assume* that all of the other cars in the parking lot are minivans (because otherwise I would have produced a more informative utterance). When the context rules out this kind of interpretation, however, the utterance is simply infelicitous. Is this kind of pragmatic infelicity generally responsible for the processing cost of negation?

Consistent with this suggestion, presenting negative information in a supportive context can mitigate some of its processing costs (Wason, 1965; Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999). When a negated feature is explicitly mentioned or inferred in preceding sentences (Lüdtke & Kaup, 2006; Orenes, Beltrán, & Santamaría, 2014), or when negation is presented within a dialogue (Dale & Duran, 2011), negative sentences tend to be processed faster relative to negative sentences presented without context. And in an ERP experiment, contextually-supported negations (e.g., "with proper equipment, scuba-diving isn't very dangerous") elicited smaller N400 responses—a marker of semantic processing costs—than unlicensed negations (e.g., "bulletproof vests aren't very dangerous"; Nieuwland & Kuperberg, 2008).

Although this previous work supports the idea that contextual expectations are the source of negation's processing cost, they do not specify the precise nature of these expectations. Our current experiment directly tests two hypotheses. First, speakers tend to produce sentences that are both relevant and informative given the context, and should be less likely to produce a sentence (negative *or* affirmative) if the context makes that sentence irrelevant or uninformative. Second, expectations about what speakers would likely say—and their match or mismatch with what the speaker in fact *does* say—are responsible for the processing costs of negation. To formalize this second hypothesis, we make use of recent probabilistic models of language comprehension, defining a listener's pragmatic expectations as the probability that a speaker would utter a statement in order to convey a particular meaning (Frank & Goodman, 2012), and using *surprisal*, an information-theoretic measure of expectation-based processing costs (Levy, 2008), to predict processing times.

The current study

In our experiment, participants viewed sets of four characters who varied in terms of appearance (i.e. hair and shirt color) as well as the presence or absence of a target feature (e.g., boys with or without apples, where apples are the "target item"). Varying the presence or absence of the target feature created five possible context conditions: $\frac{0}{4}$, $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$, $\frac{4}{4}$, where the numerator represents the number of characters in the set who e.g., have apples. In the *speaker* condition, participants produced written descriptions of one of the characters, while in the *listener* condition participants evaluated the truth value of sentences about the same pictures (see Figure 1). We predicted that speakers would be more likely to mention the presence or absence of the feature (e.g., apples) when that feature was relevant and informative. Furthermore, we predicted that listeners' processing costs would be related to the probability of a speaker using the same sentence to describe that picture.

Mentioning the target feature is only relevant when at least one other character has the target item; thus, we expected that very few (if any) participants would mention the absence of target items in the $\frac{0}{4}$ context, and that listeners would be slowest to respond to negative sentences in this context. In the $\frac{1}{4}$ context, at least one character in the set possesses target items, so mentioning the presence or absence of the target item becomes relevant in this context. We therefore predicted a sharp decrease in the surprisal of producing a true negative utterance and the reaction time to respond to a true negative utterance between the $\frac{0}{4}$ and $\frac{1}{4}$ contexts.

The informativeness of an utterances changes depending on how many other characters could be described by that utterance (e.g., saying a character "has apples" is informative when *few* other characters have apples, whereas "has no apples" is informative when *most* other characters have apples). Mentioning the absence of target items is more informative in the $\frac{1}{4}$ context compared to the $\frac{0}{4}$; however, a negative utterance in the $\frac{1}{4}$ context is still not very informative compared to describing a character's shirt color (i.e., "no apples" describes three possible characters in the $\frac{1}{4}$ context, whereas "blue shirt" only describes the referent; a participant could always produce an informative utterance by describing clothing color). We expected that the probability of producing a negative utterance would increase as the number of context characters with target items increased (i.e., in the $\frac{3}{4}$ context "no apples" is just as informative as "blue shirt", because both utterances uniquely identify the referent). For positive sentences we predicted the opposite effect of context: In the $\frac{4}{4}$ context, saying "has apples" is not very informative, because everyone has apples; in the $\frac{1}{4}$ context, however, "has apples" is very informative because the target character is the only person who has apples. We predicted a corresponding shift in reaction time for listeners, with listeners responding fastest to the sentences that were informative in context.

Method

Participants

We recruited a planned sample of 500 participants to participate in an online experiment through the Amazon's Mechanical Turk (mTurk) website; 5 participants were rejected for indicating that they were under 18 after completing the experiment. Of the remaining 495 participants, 238 were male and 255 were female, two declined to report gender, and ages ranged from 18 – 65+. We restricted participation to individuals in the US and paid 50 cents for this 10 minute study. Participants were randomly assigned to either the speaker condition (n = 242) or the listener condition (n = 253).

Stimuli

Thirty-two trial items were created in which characters were shown holding either two of the same common, recognizable objects ("target items"; e.g., two apples), or holding nothing. Within each trial, characters' shirt and hair colors also varied, providing other referential possibilities for speakers.

Each participant saw trials in which different proportions of characters were holding target items (context condition). These contexts showed $\frac{0}{4}$, $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$, or $\frac{4}{4}$ of the characters holding objects. The order of characters was shuffled on each trial, with the referent of the sentence appearing in a random position.

Participants in the speaker condition saw each image paired with an incomplete sentence (e.g. "[NAME] has ________________."). In half of the trials, the highlighted character was holding target items ("item" trials), and in half of the trials, the highlighted character was holding nothing ("nothing" trials). The experiment was fully crossed such that target characters appeared with or without target items an equal number of times in each context type.

Participants in the listener condition saw the same set of images. On each trial a sentence of the form "[NAME] [has/has no] [TARGET ITEM]" appeared. Half of the sentences were positive and half were negative (sentence type), and they were paired with pictures such that half were true and half were false (truth value), resulting in four possible trial types (true positive, true negative, false positive, and false negative). Because true positive and false negative sentences cannot occur in a $\frac{0}{4}$ context (i.e. the referent must have the target item in these trials), and true negative and false positive sentences cannot occur in a $\frac{4}{4}$ context, each trial type occurred in four possible contexts. The experiment was fully crossed, with participants receiving eight true positive, eight false positive, eight true negative and eight false negative sentences distributed equally across context types in a randomized order over the course of the study.

Procedure

Participants were first presented with a brief overview screen explaining that they would play a language game. Once participants accepted the task, they were randomly assigned to the speaker condition or the listener condition and saw more detailed instructions which explained the task and informed them that they could stop at any time.[1]

In the speaker condition, participants saw an array of four pictures on each trial: The target pictures and three context pictures presented in a random horizontal arrangement. Participants looked at these pictures for four seconds, at which point a red box appeared around one of the pictures. One second later, an incomplete sentence appeared. Participants were told that the incomplete sentence was about the picture with the box around it, and that their job was to complete the sentence (by typing into a small text box) using only a few words. Participants were explicitly discouraged from referring to the character's position in the array of pictures.

In the listener condition, participants first saw eight positive sentence practice trials with feedback about incorrect responses before beginning the test trials. In each test trial, participants saw the same arrays of pictures shown to participants in the speaker condition for four seconds, at which point a red box appeared around one of the pictures. One second later, a sentence about that picture appeared. Participants were told to read the sentence and respond as quickly and accurately as possible with a judgment of whether it was true or false when applied to the highlighted picture. We recorded reaction times for each trial, measured as the time from when the sentence was presented to the moment when the response was made.

Data Processing

We excluded 16 participants who did not list English as their native language and two participants from the listener condition for having an overall accuracy below 80%.[2] This left a total of 477 participants for analysis (244 in listener condition, 233 in speaker

[1] The speaker and listener conditions of the experiment can be viewed at `http://anordmey.github.io/negatron/experiments/speakers/negatron.html` and `http://anordmey.github.io/negatron/experiments/listeners/negatron.html`, respectively.

[2] Overall accuracy on this task was very high, with participants responding correctly on 98% of true positive trials, 93% of true negative trials, 96% of false positive trials, and 97% of false negative trials.

condition).

In the speaker condition, affirmative responses labeling the target feature were coded as “positive” (e.g., “apples,” “two apples,” “red apples,” etc.). Responses negating the target feature (e.g., “no apples”) were coded as “negative.” All other responses (e.g. descriptions of the characters’ clothing or hair color, as well as other types of positive or negative utterances) were coded as “other.” Codes were hand-checked to ensure that label synonyms or spelling errors were coded correctly.

In the listener condition, we excluded trials with RTs greater than 3 standard deviations from the log-transformed mean, a criterion established in our previous experiments (Nordmeyer & Frank, 2014).

To predict responding in the listener condition, we used productions from the speaker condition. We calculated the proportion of positive sentences describing characters who possessed target items, and the proportion of negative sentences describing characters with nothing, creating probability distributions for true positive and true negative utterances in each context. We then used this distribution to calculate the surprisal of hearing a true positive or true negative sentence for each context. Surprisal is an information-theoretic measure of the amount of information carried by an event (in this case, the amount of information conveyed by a sentence); in prior work on sentence comprehension it has been used successfully as a linking hypothesis between production probabilities and reaction times (Levy, 2008). Surprisal (or “self-information” I) for a sentence s is defined as

$$I(s) = -\log(P(s)). \tag{1}$$

Results

Speaker Condition

Participants were more likely to produce true positive utterances about target items than they were to produce true negative utterances about the *absence* of target items. As the number of characters in the context with target items increased, however, participants became less likely to produce true positive utterances about referents with target items, and more likely to produce true negative utterances about characters without target items (see Figure 2). That is, the production of both negative *and* positive sentences was influenced by the surrounding context. These findings support our hypothesis that speakers will produce informative utterances (i.e. produce sentences that are maximally effective at identifying the referent), because true positive utterances about the presence of target items are more informative when none of the other characters have target items (e.g. the $\frac{1}{4}$ context), whereas true negative utterances about the absence of target items are more informative when the other characters *do* have target items (e.g. the $\frac{3}{4}$ context)

Consistent with our hypothesis that speakers will produce negative sentences that are relevant (i.e. negating a salient feature of the context), negative sentences were almost never produced on "nothing" trials in the $\frac{0}{4}$ context. That is, "no apples" was a very rare production on trials where there weren't any apples in the context; participants in this condition tended to produce affirmative sentences describing other features of the referent, such as their clothing or hair color. Participants only consistently produced negative utterances in the $\frac{1}{4}$ context, where at least one other character has target items, thus making the target items a relevant feature to discuss.

To evaluate the reliability of these patterns, we fit two separate binomial mixed-effects models: 1) a model fit only to negative utterances, and 2) a model fit only to positive utterances. To test the effect of informativeness, we coded context as numeric, e.g. the proportion of characters in the context with target items. To test the effect of relevance, we created a dummy code to separately test for the effects of the $\frac{0}{4}$ context compared to all of the other contexts (i.e., the relevant contexts).[3]

[3]All mixed-effects models used the maximal convergent random effects structure (Barr, Levy, Scheepers, & Tily, 2013) and were fit using the lme4 package version 1.1-7 in R version 3.1.2. Raw data and analysis code can be found at `https://github.com/anordmey/negatron`.

For the first model described above, we tested the effect of context on the probability of producing a true negative utterance on “nothing” trials.[4] We found a significant effect of the relevant contexts, indicating that participants were less likely to produce negative sentences in the $\frac{0}{4}$ context compared to the other contexts ($\beta = -1.19$, $p = .015$). This finding suggests that relevance has a significant effect on the production of true negative sentences. We also found a significant positive linear effect of context, with the probability of producing a negative sentence increasing as the proportion of characters with target items increases ($\beta = 7.30$, $p < .001$), indicating a significant effect of informativeness on the production of negative sentences.

The second model tested the effect of context on the probability of producing a true positive utterance on “item” trials.[5] We found a significant *negative* linear effect of context ($\beta = -5.5$, $p < .001$), indicating that the probability of producing a positive utterance decreases as the proportion of characters with target items increases. The findings of both models together suggest that the probability of producing both true positive and true negative utterances is influenced by the informativeness of that utterance in context.

Figure 3 shows the surprisal of true positive and true negative sentences. Consistent with our predictions, surprisal decreases for true negative sentences and increases for true positive utterances as the number of characters with target items increases, indicating a significant effect of informativeness on the production of both positive and negative sentences. The steep gap in surprisal between the $\frac{0}{4}$ and the $\frac{1}{4}$ contexts is driven by the fact that almost no participants produced negation in the $\frac{0}{4}$ context, where a negative utterance is not relevant. Because the probability is close to zero, the surprisal for these utterances is much higher than the surprisal of an utterance that is even slightly more probable.

In our analyses we coded utterances such as “Bob has nothing” as “other” rather

[4]The model specification was as follows: `negation ~ relevant context + numeric context + (1 | subject) + (1 | item)`.

[5]The model specification was as follows: `positive ~ numeric context + (1 | subject) + (1 | item)`.

than "negative" because we wanted to examine speaker utterances that were as close as possible to the utterances in the listener condition. Furthermore, our definition of relevance is specific to the negation of the target item (e.g., negating "apples", not just producing any negation). In two exploratory analyses, we examined how these coding decisions influenced our results. Consistent with our predictions, if we include other negations of the target item in our analysis (e.g., "not apples", "without apples", "zero apples", etc.), the results of the speaker condition are the same as those reported above (i.e., significant effects of relevance and informativeness). If we include instances of "nothing", however, the effect of informativeness remains highly significant ($\beta = 6.55$, $p < .001$), but the effect of relevance is no longer significant ($\beta = -0.21$, $p = 0.50$). That is, "has no apples" is not a very relevant utterance in a context where no one else has apples, but "has nothing" is, perhaps, a relevant thing to say in the context of an experiment where people sometimes have objects and sometimes do not.

Listener Condition

Participants were fastest to respond to true positive sentences, and slowest to respond to true negative sentences, replicating previous findings (H. Clark & Chase, 1972). Listeners' responses to true negative sentences mirrored the surprisal of true negative sentences in the speaker condition, with participants responding slowest to true negatives in the $\frac{0}{4}$ context. The same pattern was seen in response to false positive sentences, suggesting that listeners expect speakers to describe relevant features of the context even when the sentence is false (see Figure 4).

We fit a linear mixed-effects model to examine the interaction between sentence type (positive or negative), truth value (true or false), and context as predictors of reaction time.[6] All model coefficients are shown in Table 1. In addition to main effects of sentence type ($\beta = -205$, $p < .001$) and truth value ($\beta = -384$, $p < .001$), there was an interaction

[6]The model specification was as follows: `RT ~ sentence × truth × context + (sentence | subject) + (sentence | item)`.

such that true positive sentences elicited the fastest responses and true negative sentences elicited the slowest responses ($\beta = -663$, $p < .001$). The model showed a significant negative linear effect of context, with reaction times decreasing as the proportion of characters with target items increased ($\beta = -341$, $p < .001$). A significant three-way interaction between sentence type, truth value, and context indicates that this pattern was driven primarily by responses to true negative sentences, with context having the most pronounced effect on true negative utterances ($\beta = -901$, $p < .001$).

To explore the separate effects of relevance (i.e., the effect of the $\frac{0}{4}$ context compared to the others) and informativeness (i.e. the linear effect of context) on responses to true utterances, we fit two separate models to reaction times in response to true positive and true negative utterances.[7] We found a significant effect of relevance on reaction times for negative utterances, with the $\frac{0}{4}$ context producing significantly slower reaction times compared to the relevant contexts ($\beta = 307$, $p < .001$). We did not find a significant linear effect of context above and beyond the effect of the $\frac{0}{4}$ context ($\beta = -41$, $p = .58$). We did, however, find a significant positive linear effect of context on the the reaction time to respond to positive sentences ($\beta = 111$, $p = .02$), indicating a significant effect of informativeness on RTs to positive sentences, but not negative sentences.

Predicting Listeners with Speakers

To test our primary hypothesis that processing times are a function of listeners' expectations about what a speaker will say, we regressed the mean reaction time in response to true positive and negative utterances in each condition against the surprisal for the same utterances (Figure 5). There was a significant positive relationship between surprisal and reaction time for true negative sentences, $r^2 = .93$, $p < .001$, supporting our

[7] Model specification for true negative model: `rt ~ relevance context + numeric context + (relevant context + numeric context | subject) + (relevant context + numeric context | item)`; model specification for true positive model: `rt ~ numeric context + (numeric context | subject) + (numeric context | item)`.

prediction that the effects of context on reaction time reflect differences in how speakers would describe the same stimuli. This relationship between surprisal and reaction time for true negative sentences holds even with the outlying $\frac{0}{4}$ context removed from analysis ($r^2 = .71$, $p = .017$).

Much like other psycholinguistic experiments, our tasks take place in a restricted world where the set of possible referents (and by extension, the set of possible utterances that a speaker is likely to produce) is limited. In real world contexts, a speaker could produce many different utterances to describe many different referents at any given moment. Our goal in these experiments is to explore whether Neo-Gricean pragmatic principles can at minimum explain how people communicate in these restricted worlds. Our theory would predict that when people produce negative utterances in the "real world" they are *usually* producing them in highly informative and relevant contexts; thus, we would predict that people have less difficulty processing negative sentences in natural conversation than they do in constrained psycholinguistic experiments.

We do have some data to explore whether the relationship between speaker surprisal and listener reaction time is influenced by the availability of other possible utterances. In a previous, unpublished version of this study, participants engaged in a version of these tasks where all four characters were identical except for the presence or absence of target items.[8] Thus, in the previous version, descriptions of clothing color were always relatively uninformative (because they were true of all characters in a given trial). A negative utterance on "nothing" trials was therefore at least as informative as describing shirt color in the $\frac{1}{4}$, $\frac{2}{4}$, and $\frac{3}{4}$ contexts. Contrast this with the current study, where describing the color of a characters shirt is *always* an informative utterance; the informativeness of a true negative utterance is only comparable to these competing utterances in the $\frac{3}{4}$ context. The effect of the informativeness of these competing utterances in the current study is reflected in the raw probabilities for the speaker data (Figure 2), where there is a jump in the

[8] Raw data and analysis code are available at `https://github.com/anordmey/negatron`.

probability of a true negative utterance between the $\frac{2}{4}$ and the $\frac{3}{4}$ context. In the previous version, this sharp increase occurred between the $\frac{0}{4}$ and $\frac{1}{4}$ contexts – again, the point where producing a true positive utterance becomes as informative as producing an utterance about e.g., shirt color.

In both the previous study and the current study, the *surprisal* of producing a negative utterance is significantly higher in the $\frac{0}{4}$ than in any other context, because the probability of producing a true negative utterance in this context is close to zero. Furthermore, in the listener condition of both versions of the experiment the $\frac{0}{4}$ context yields much slower reaction times than any other context. Critically, replicating both the direction and magnitude of the findings reported here, in the previous study, speaker surprisal and lister reaction time were strongly related, $r^2 = .89$, $p < .001$. Thus, the relationship between speaker surprisal and listener reaction time appears to be robust to changes in the relative informativeness of competing utterances.

General Discussion

What makes negation hard to process? While previous work has proposed that processing negative elements is especially difficult because of features intrinsic to negation, our work suggests that the same general pragmatic mechanisms that govern positive sentences are responsible for much of the difficulty associated with negative sentences. Negative sentences presented without context are uninformative and irrelevant; thus, they are unlikely to be produced by speakers. In turn, listeners respond to unlikely utterances with increased processing times. In contexts where negation is more informative and relevant, processing costs are lower. Overall, this evidence supports a Gricean interpretation of negation processing, with pragmatic principles playing a role in the processing of both positive and negative sentences.

While previous work has shown that contextual factors facilitate the processing of negation (Wason, 1965; Nieuwland & Kuperberg, 2008; Dale & Duran, 2011; Orenes et al., 2014), our findings here go further. First, by using actual language productions as the predictor of processing difficulty, our work strongly implicates specifically pragmatic factors. Because the field of pragmatics is concerned with language *use*, demonstrating a relationship between actual speaker productions and listener processing time is critical to the argument that pragmatic factors are responsible for the processing cost of negation. To our knowledge none of the past work on the effect of context on negation has demonstrated this relationship. Second, rather than treating pragmatics as a black box, we show that two different components—informativeness and relevance—each contribute to the relative (un-)likelihood of hearing a negation. Speakers were unlikely to mention the absence of a feature unless it was a relevant feature in the context, and were more likely to mention the presence or absence of features when mentioning those features would help uniquely identify the referent. The fact that informativeness played a role in the production of both positive and negative utterances supports our argument that these pragmatic pressures are general, rather than specific to negation.

Listeners' reaction times were highly correlated with the surprisal of a speaker producing the same utterances in context. For true negative sentences, however, this effect was driven primarily by the effect of relevance (i.e., the difference between the $\frac{0}{4}$ and all other contexts). Why didn't informativeness play a role in listeners' response times? One possibility is that only relevance plays a role in forming listeners' expectations about what a speaker will say. Another possibility is that the effect of informativeness is small compared to the effect of relevance, and that the reaction time measure is too noisy to identify this effect. Although our results cannot disentangle these two possibilities, the fact that we did find a significant effect of informativeness on listeners' responses to true positive sentences suggests that informativeness does play some role in listeners' expectations.

We use surprisal, an information-theoretic measure of processing difficulty, as the linking hypothesis between speaker probabilities and reaction times (Levy, 2008). Although this link has substantial support in the realm of syntactic processing (Demberg &

Keller, 2008; Boston, Hale, Kliegl, Patil, & Vasishth, 2008), to our knowledge, our findings are the first example of using surprisal over sentence-level pragmatic expectations, rather than word-level syntactic expectations. This success suggests that, in concert with the appropriate predictive models, surprisal theory could be productively applied to the prediction of processing difficulty beyond the level of syntax.

Our analysis is intended as a computational level analysis (Marr, 1982). We are not committed to a specific account of how pragmatic inferences are represented. Our results don't tell us whether listeners actually simulate a speaker when they form expectations about what a speaker will say, or doing something that approximates such a simulation. Our goal is to show at least that the weaker of these two possibilities is true; that is, listeners are forming expectations about what a speaker will say, even though we don't know precisely what form these representations take. Several mechanistic theories of pragmatics have been proposed, such as an in-the-moment alignment of linguistic representations between speakers and listeners (Pickering & Garrod, 2004), or cached expectations about how speakers use language leading to preferred interpretations by listeners (Levinson, 2000). Either of these possibilities, or any number of others, could be consistent with our data.

Although we focus here on negation, our findings have implications for sentence processing more generally. Debates about the effects of pragmatics on linguistic processing exist in other domains, such as the processing of scalar implicatures (the pragmatic inference that e.g., "some" implicates "some but not all"; Huang & Snedeker, 2009, 2011; Grodner, Klein, Carbary, & Tanenhaus, 2010). Tomlinson, Bailey, and Bott (2013) provide an informative comparison between scalar implicature and negation, presenting mouse-tracking trajectories for each. Their negation data show the same pattern of processing difficulties we observe, and critically, their data on the processing of underinformative "some" utterances look almost identical. We hypothesize that, in both cases, participants' processing difficulty is a function of the violation of their pragmatic expectations.

Our findings here suggest that a large part of the processing difficulties of negative sentences arise from the relative pragmatic felicity of negation in context. They do not rule out the possibility that there is some co st to processing an additional logical element, but this processing cost would have to be quite small with respect to the magnitude of the effects we observed. This finding leads us to the following conclusion: When logical words are used in a communicative context, we have no difficulty understanding them.

Author Contributions

Both authors developed the study concept and contributed to the study design. Data collection was conducted by A. E. Nordmeyer. A. E. Nordmeyer performed the data analysis and interpretation under the supervision of M.C. Frank. Both authors contributed to the development of the manuscript and approved the final version of the manuscript for submission.

Acknowledgments

This work supported by the NSF GRFP to AEN and ONR N00014-13-1-0287.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Boston, M., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *The Mind Research Repository (beta)*.

Carpenter, P., & Just, M. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*, 45–73.

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

Clark, H. H. (1976). Semantics and comprehension.

Dale, R., & Duran, N. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*, 983–996.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*, 193–210.

Ferguson, H., & Sanford, A. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, *58*, 609-626.

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.

Glenberg, A., Robertson, D., Jansen, J., & Johnson-Glenberg, M. (1999). Not propositions. *Journal of Cognitive Systems Research*, *1*, 19–33.

Grice, H. (1975). Logic and conversation. *1975*, 41–58.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics*, *38*, 1015–1032.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, D.C.: Georgetown University Press.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*, 376–415.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*, 1161–1172.

Just, M., & Carpenter, P. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*, 244–253.

Just, M., & Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.

Kaup, B., Ludtke, J., & Zwaan, R. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*, 1033–1050.

Kaup, B., & Zwaan, R. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 439–446.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lüdtke, J., Friedrich, C., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *The Journal of Cognitive Neuroscience*, *20*, 1355–1370.

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1735–1740).

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York, NY: Henry Holt and co. Inc.

Moeschler, J. (1992). The pragmatic aspects of linguistic negation: Speech act, argumentation and pragmatic inference. *Argumentation*, *6*, 51–76.

Moxey, L. M. (2006). Effects of what is expected on the focussing properties of quantifiers: A test of the presupposition-denial account. *Journal of Memory and Language*, *55*, 422–439.

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.*

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36–45.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*, 169–190.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford, UK: Blackwell Publishing.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*, *63*, 2305–2312.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*, 18–35.

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

Table 1

Coefficient estimates from a mixed-effects model predicting listeners' reaction times in response to sentences in different contexts.

	Coefficient	Std. err.	t
Intercept	1598	41	39.16
Sentence (Negative)	-205	38	-5.36
Truth (True)	-384	36	-10.57
Context	-341	42	-8.14
Sentence × Truth	663	54	12.16
Sentence × Context	377	59	6.40
Truth × Context	454	59	7.71
Sentence × Truth × Context	-901	83	-10.79

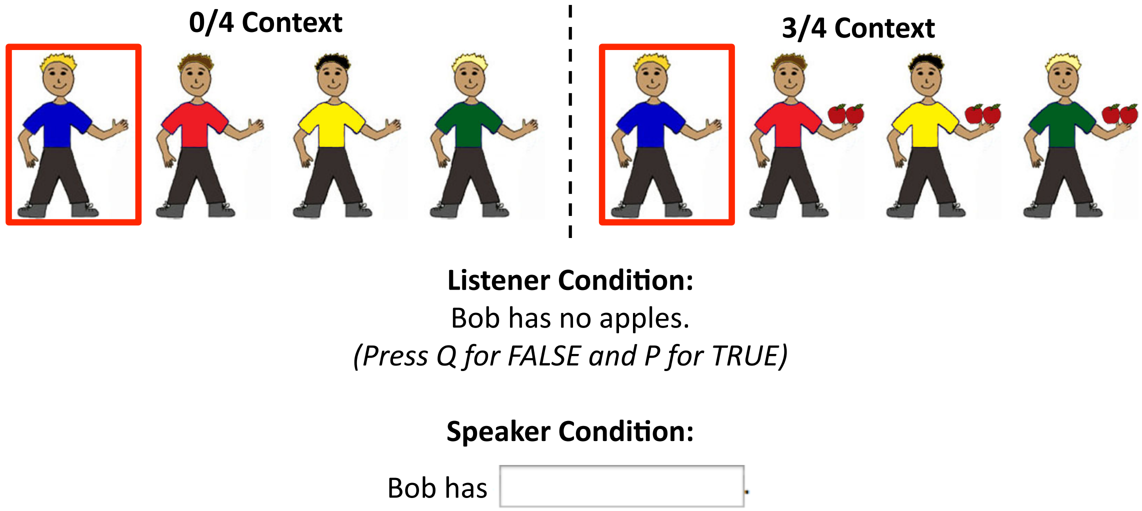

Figure 1. An example of a true negative trial with a 0/4 context (left) and a 3/4 context (right). The sentence "Bob has no apples" in the 0/4 context is both uninformative (because the sentence is true of all of the characters) and irrelevant (because apples are not present in the context), whereas the same sentence in the 3/4 context is both informative and relevant.

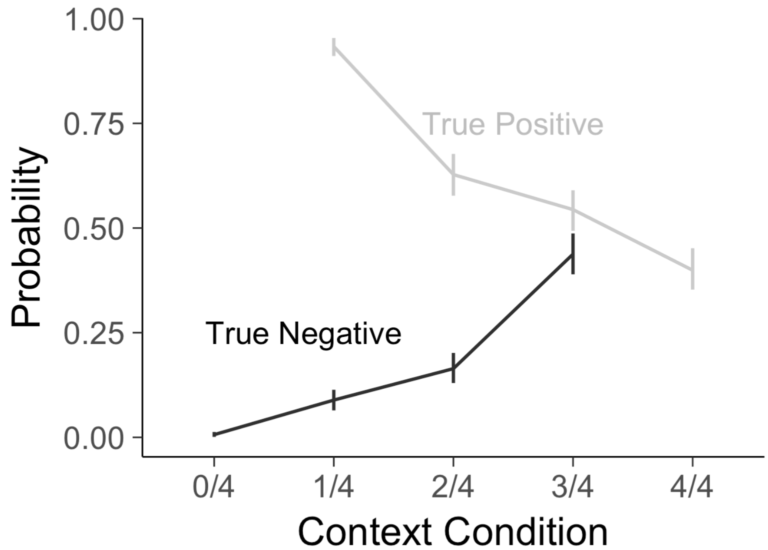

Figure 2. Probability of producing negative sentences on “nothing” trials (i.e., true negatives) and positive sentences on “item” trials (i.e., true positives) across different contexts. Negative sentences are shown in black, and positive sentences in grey. The context is notated by a fraction representing the number of characters in the context who held target items. Error bars show 95% confidence intervals computed by non-parametric bootstrapping.

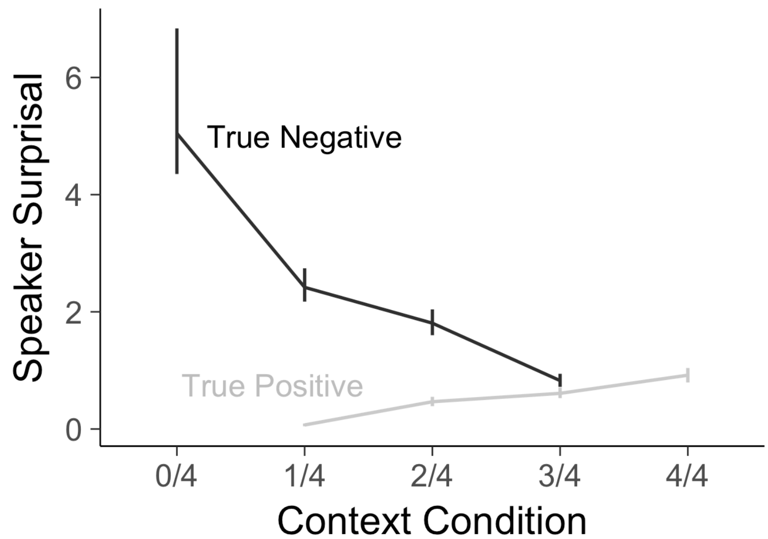

Figure 3. Surprisal for true positive and true negative sentences across different contexts. Negative sentences are shown in black, and positive sentences in grey. The context is notated by a fraction representing the number of characters in the context who held target items. Error bars show 95% confidence intervals computed by non-parametric bootstrapping.

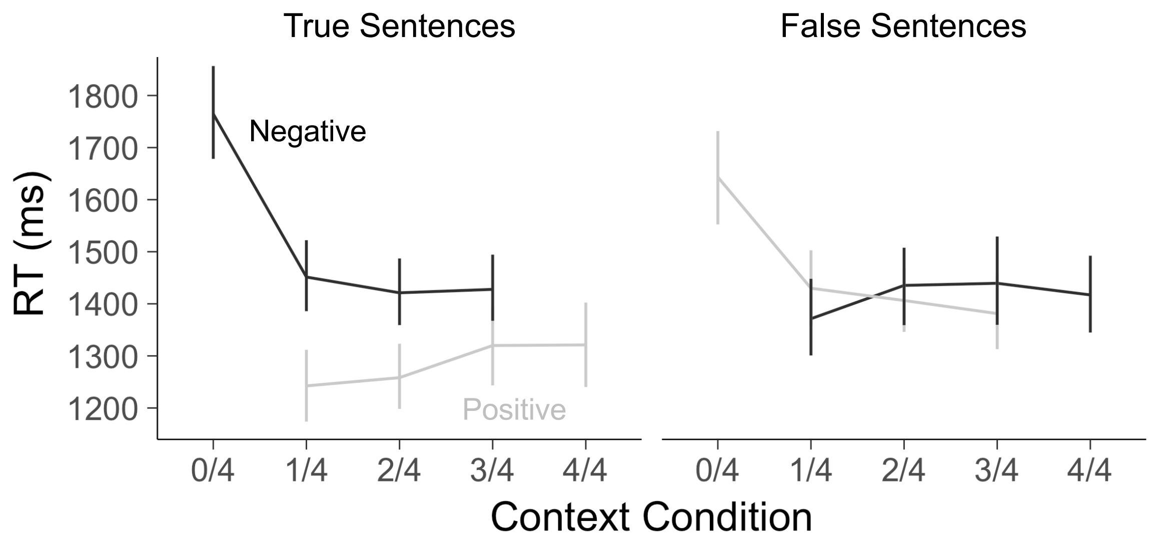

Figure 4. Reaction times for each trial type across different conditions. Responses to true sentences are shown on the left, and false sentences are shown on the right. Negative sentences are shown in black, and positive sentences in grey. The context is notated by a fraction representing the number of characters in the context who held target items. Error bars show 95% confidence intervals computed by non-parametric bootstrap.

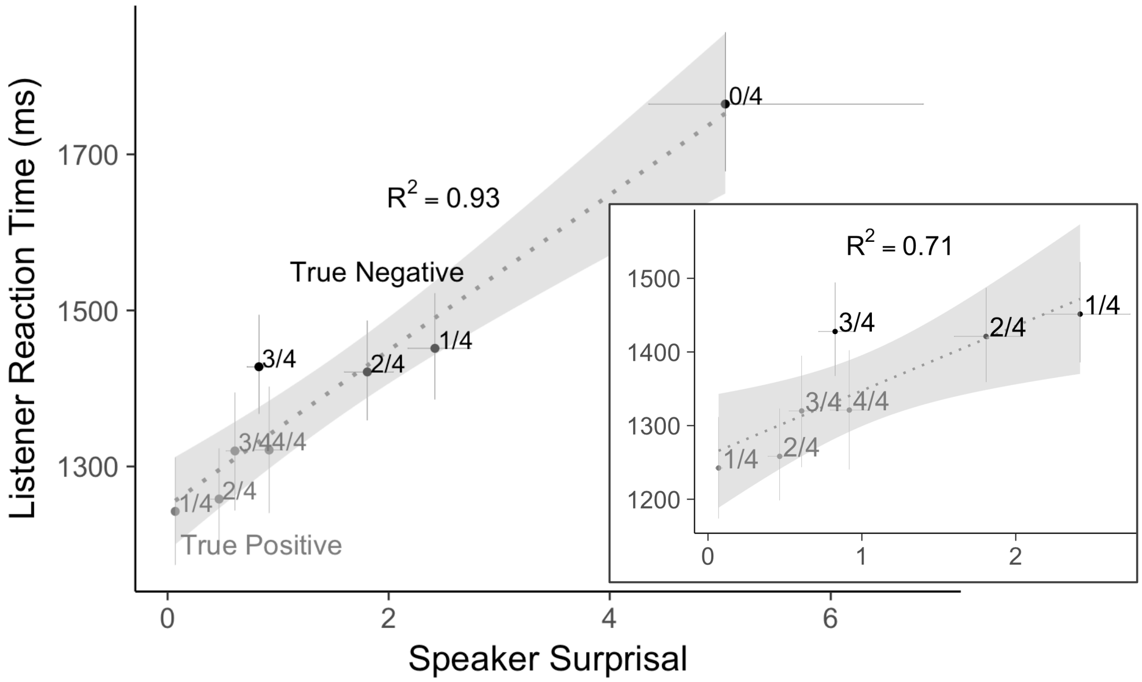

Figure 5. Reaction times in the listener condition plotted by surprisal in the speaker condition. Each point represents a measurement for sentence type and context. Negative sentences are shown in black, and positive sentences in grey. Error bars on the horizontal and vertical axes represent 95% confidence intervals on their respective measures. The gray band shows the linear regression and 95% confidence region for all conditions. The inset graph zooms in on the linear regression and 95% confidence region for data excluding the outlying 0/4 condition.