Replication of “The Cognitive Dynamics of Negated Sentence Verification” by Dale & Duran (2011, *Cognitive Science*)

Ann Nordmeyer
anordmey@stanford.edu

Introduction

This project replicates studies 1 and 3 of “The Cognitive Dynamics of Negated Sentence Verification”, Dale & Duran (2011). The original study 1 used mouse-tracking to examine the differences in processing positive and negative sentences. They found that negative sentences elicit more discrete mouse movements (according to two measures of discreteness) while evaluating whether the sentence is true or false. Study 3 examined how this finding changes when the sentences are presented in a pragmatically licensed context. According to one of their measures (x-flips, i.e. the number of times the mouse trajectory changed along the x-axis), there was no difference in discreteness between positive and negative sentences when presented in a supportive context. According to their other measure (acceleration components, i.e. the number of times the mouse trajectory sped up or slowed down), responses to negative sentences were still more discrete than responses to positive sentences, but discreteness was significantly reduced from study 1.

Methods

Power Analysis

Original effect size was unknown, but a power analysis via simulation indicates that the original experiment 1 had 95% power with N=60 participants. Power for experiment 3 was not calculated since the null hypothesis was confirmed (i.e. there was no effect of sentence type on reaction time).

Planned Sample

Planned sample size was 60 participants for each condition (context v. no context) for a total of 120 participants. Only MTurk participants residing in the United States were eligible to participate in the study.

Materials

Experiment 1:
From original:
“The interface was programmed using Adobe Flash, permitting extraction of x,y-pixel coordinates at a sampling rate of approximately 40 Hz (see Fig. 1). The software filled the window of the participant’s browser. We created 12 sentences to serve as a basis for true ⁄ false statements. Each of these sentences could be a true or false statement in both negated and nonnegated forms.1 As examples, ‘‘elephants are not small’’ is an example of a

negated true statement, made false by changing the adjective: ‘‘elephants are not large’’ (a false, negated statement). Another example stimulus was ‘‘cars have wings,’’ a false statement without negation, but true when negated: ‘‘cars have no wings.’’ Importantly, each subject saw an equal number of these four types, but with each sentence randomly assigned to one type.”

Experiment 3:
Experiment 3 used the same interface as experiment 1, but with participants judging sentences as “sensible” or “nonsense”. Sentences were embedded in a context where the statement to be evaluated was a sentence from an adult spoken to a child.
From original:
“Experiment 3...embedded [sentences] in an enhanced pragmatic context, which was designed as a statement from an adult to a child. For example, ‘‘You want to lift an elephant?’’ is a statement that may set up anticipation for plausible denial from an adult who is correcting a child, next stating, ‘‘but elephants are not small.’’ The items were then converted into a quote, such as: ‘“You want to lift an elephant?’ the mother said to her child, ‘but elephants are not small’.’’ This was done for each item, and participants were told to judge whether the statements by adults were sensible or nonsense. This experiment thus embeds the items in a strong pragmatic context and changed the nature of the response: Participants now judged the sensibility of adult statements to the child. This increased the pragmatic license of negation, and participants were
instructed to focus on the whole statement and judge it for sensibility. Crucially, negation occurred at exactly the same distance from the final word cuing a participant’s response.”

This was followed precisely, with the exception that the interface was programmed using javascript, and was presented within the Amazon Mechanical Turk website, so the software filled the frame but not the participant’s entire browser window. Frame height was fixed at 700 px. Exact sentences used and instructions given to participants were requested from the original researchers. Paradigm can be viewed at the following link: http://www.stanford.edu/~anordmey/psych254/rep_project/negmouse.html. See Figure 1 for an example of the interface.

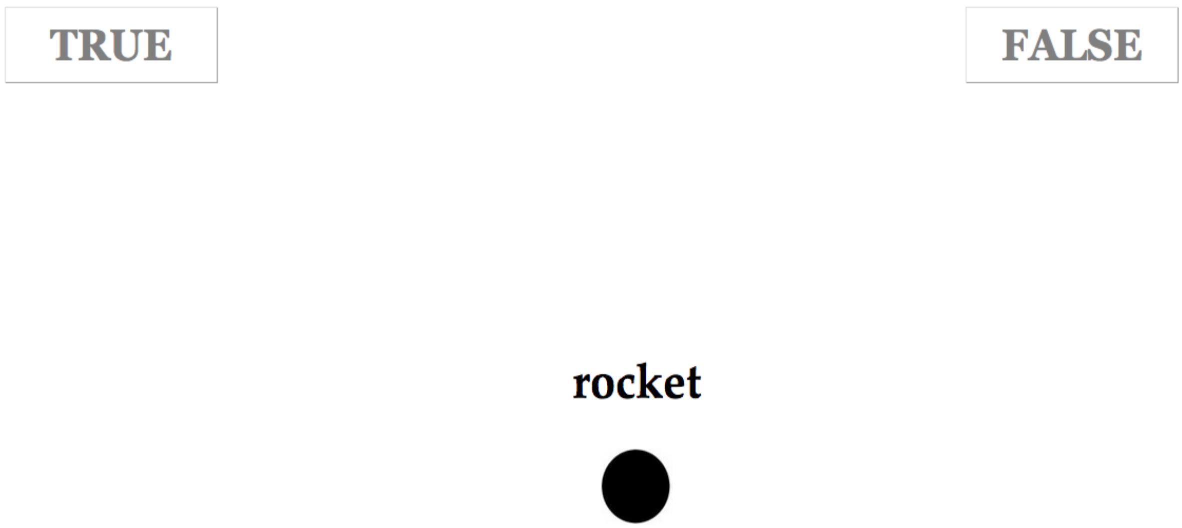

Figure 1: A screenshot of the experiment interface

Procedure

From original:

"Sentences were presented one word at a time in a self-paced reading format. After initial instructions, participants saw TRUE and FALSE boxes at the top-left or top-right of the interface (placement was randomized across participants, but consistent within). They then clicked a circle at the bottom center of the interface, revealing a word immediately above this circle. They clicked until the final word appeared, at which point the circle disappeared, and a response was made to either TRUE or FALSE as appropriate. Participants saw all 12 sentences, with each sentence randomly assigned to negated ⁄ not negated and true ⁄ false, counterbalanced to ensure participants saw an equal number of each kind. The x,y-pixel trajectory was saved, and it consisted of their movement from their last word click (from the bottom center) to their TRUE⁄FALSE click (to the top-left or top-right). Importantly, the negation in these sentences appears several words before a response is made. This ensured that any results we obtained were not purely stimulus-induced, because the negating element (most often ''not'') always occurred 1–2 words before the end of a sentence."

This procedure was followed precisely. Sentences were presented in a random order for each participant, as was done in the original experiment. In experiment 3, the word "TRUE" was replaced with "sensible" and the word "FALSE" wa replaced with "nonsense". This was the case in the original experiment as well. The experiment took 5-10 minutes to complete, and participants were paid 50 cents for their participation.

Analysis Plan

From original:

"We chose two trajectory measures that imply more discrete changes in the mousemovement trajectory, both exemplified in Fig. 1. First, we calculated x-flips (Dale et al., 2008), which is a count-based score of the number of times the mouse cursor goes back and forth along the x-axis (i.e., the axis of decision):

$$x\text{-flips} = \sum H[-(\Delta x_t - \Delta x_{t-1})(\Delta x_{t-1} - \Delta x_{t-2})]$$

x_t represents the x-axis pixel coordinate at time t. H represents the Heaviside function, a threshold function that will return 1 when there is a flip of directionality by taking the product of -1 and three-step comparisons of directional change along the x-axis. If x is increasing from t-1 to t (+ change), and decreasing from t-2 to t-1 (- change), the product of these differences will always be -1, thus the negative sign produces a positive product, and H functions to produce an output of 1. By conducting this three-window analysis across the trajectory (indicated by the summation sign), we obtain a count of the number of switches.

Second, we used what we will here term ‘‘acceleration components’’ (AC). Similar measures have been used in studies on error correction in low-level motor control (Fishbach, Roy, Bastianen, Miller, & Houk, 2005), and it reflects the number of times the trajectory accelerates ⁄ decelerates during the response:

$$\mathrm{AC} = \left(\sum H[-(a_t - a_{t-1})(a_{t-1} - a_{t-2})]\right) - 1$$

AC is defined in the same way as x-flips, but at above reflects the acceleration at time t. When acceleration changes direction (going from positive acceleration to negative acceleration) there is more complexity in the programmed movement (Fishbach et al., 2005; also see Wojnowicz et al., 2009 for analysis of acceleration complexity). The subtraction of 1 is to factor out the standard change in acceleration that is seen in a basic movement (even the simplest, straight movement will have one instance of positive to negative acceleration). The measure of x-flips reflects a complexity in the direction of movement—a spatial shift occurring during action dynamics. AC complements x-flips as discreteness may be present in the unfolding movement without necessarily a change in direction. For example, a participant may indicate a temporal fluctuation in their movement but not shift direction.
Images exemplifying both measures are shown in Fig. 1. Thus, the two measures offer complementary ways of detecting abrupt changes in unfolding cognitive dynamics.

The above review of the sentence verification literature suggests two straightforward predictions from these measures. If negation changes cognitive processing in ways that go beyond just the time required for a decision to be made (reaction time), we should see increased discreteness or complexity of unfolding action execution. Specifically, if negation produces abrupt shifts in cognitive dynamics, then x-flips should increase in count for sentences with negation. This should also occur for AC, because the integration may require a temporary break from the smooth dynamics of a previous interpretation, indexed by the arm’s movement.

To analyze these measures, we used a linear mixed-effects analysis in the way described by Baayen, Davidson, and Bates (2008), using a 2 (negation vs. no negation) by 2 (true vs. false) full factorial, repeated-measures model. Subject and sentence topic were used as random factors. This simultaneously controls for subject- and item-derived effects.”

These two measures of discreteness will both be used, and a similar maximal mixed-effects analysis will be used to analyze the measures described here. As in the original experiment, participants who respond correctly less than 80% of the time will be discarded from analysis. In addition, trials with reaction times 3 standard deviations away from the mean reaction time will also be discarded, as was done in the original experiment. Reading times will also be collected in this experiment, though they were not analyzed in the original experiment.

Differences from Original Study

The primary known difference, at this point, is that the interface used will be programmed using javascript instead of Adobe Flash, and will be presented within a smaller window (with fixed height of 700 px) on the MTurk page, rather than filling the participant's entire browser window. These changes are not anticipated to have an effect on the outcome of the studies. The original sample was also collected using Mechanical Turk, and the precise sentences and instructions used were requested from the original researchers.

(Post Data Collection) Methods Addendum

Actual Sample

120 participants were recruited and participated through Amazon Mechanical Turk, 60 in the "no context" condition (i.e. experiment 1) and 60 in the "context" condition (i.e. experiment 3). Participants were between 18 and 65 years of age, with most participants falling between the ages of 18 and 35 (n=97) and the majority between 25 and 35 years of age (n=56). 47 participants were female and 73 were male. Three participants were excluded for listing a language other than English as their native language. Two additional participants were excluded due to the fact that one did not use a traditional mouse or trackpad and no trajectory information was collected, and the other had no trajectory information for unknown reasons. 14 participants were excluded for answering less than 80% of trials correctly. This left a final sample of N=102, with 55 participants in the No Context condition and 46 participants in the Context condition.

Differences from pre-data collection methods plan

None

Results

Data preparation

Nineteen participants were excluded from analysis (Non-English speaker (n=3), no trajectory information (n=2), under 80% correct (n=14)), leaving a final sample of 101 participants (no context: n = 55; context: n = 46).

In addition, trials in which participants answered incorrectly were excluded from analysis. This resulted in the removal of 32 trials. Furthermore, trials with reaction times three standard deviations greater than the mean reaction time for each condition were removed from analysis. This resulted in 2 trials being removed from analysis.

X- and y- coordinates were transformed such that y-coordinates equal to 0 corresponded to the bottom of the experiment window, and x-coordinates equal to 0 corresponded to the midline. This was done to make x- and y-coordinates easier to interpret, but does not alter the analysis in any way.

X-flips and acceleration components were calculated for each trial. X-flips were calculated by taking the derivative of the vector of x-coordinates and counting the number of times there was a change from negative to positive or positive to negative. Acceleration components were calculated by determining the acceleration in each 25 ms time point, and counting the number of times there was a change from negative acceleration to positive acceleration or vice versa.

Confirmatory analysis

The original paper (Dale & Duran, 2011) used two measures to analyze discreteness of the mouse trajectory. These were x-flips (the number of times the trajectory changed along the x-axis), and acceleration components (the number of times the acceleration sped up or slowed down).

X-flips:

In Dale & Duran (2011) Experiment 1 (no context), a significant effect of sentence type (positive v. negative) was found, such that negative sentences produced more x-flips than positive sentences. Additionally, a significant interaction between sentence type and truth value was found, such that true positive sentences showed the fewest x-flips and true negative sentences showed the most x-flips. In Experiment 3 (context), no effect of sentence type or truth value was found, and the interaction between sentence type and truth value was not significant. The original means are plotted in Figure 2.

Replication means can be seen in Table 1. The overall pattern of means was very similar to that seen in Dale & Duran (2011), as can be seen in Figure 2. Mixed effects models with random effects of subject and item were tested separately for the no context and the context condition[1]. In the no context condition, a marginal effect of negation was found, but the interaction between sentence type and truth value was not significant. In the context condition, there were no significant main effects or interactions. Results of these models can be seen in Table 2.

---

[1] For the no context condition, the maximal random effects structure was used: (sentence.type*truth.val | subid) + (sentence.type*truth.val | item). However, for the context condition, the maximal model would not converge and therefore only additive random effects of items were included: (sentence.type*truth.val | subid) + (sentence.type+truth.val | item)

Noting that the distribution of x-flips resembled a poisson distribution, with most trials having zero x-flips, I attempted an additional analysis of the no context condition in which I treated the x-flips results as binomial, such that x-flips = 0 were given a value of 0 and x-flips > 0 were given a value of 1. I then tested a binomial mixed effects model with random effects of subject. However, the binomial model did not yield any significant effects.

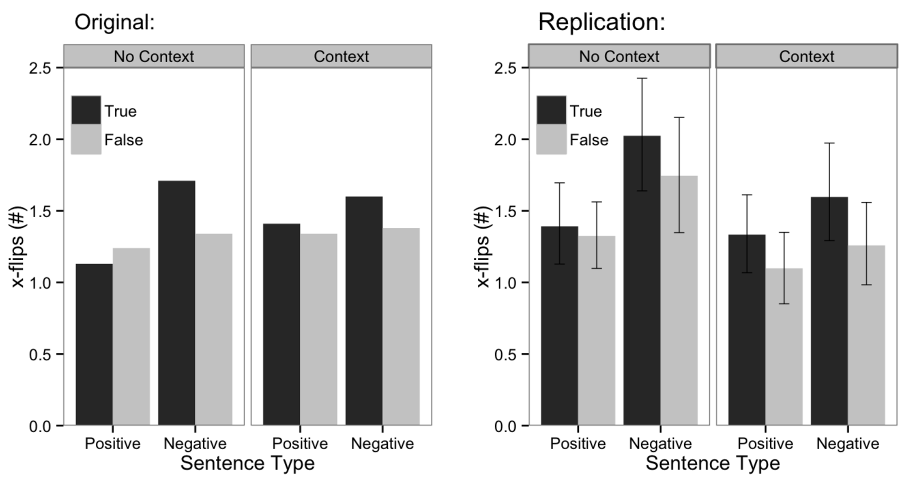

Figure 2: Original means (left) and replication means (right) for number of x-flips. Error bars for replication data are bootstrapped 95% confidence intervals.

	Exp. 1 / No Context		Exp. 3 / Context	
	Original	Replication	Original	Replication
True Positive	1.13	1.39	1.41	1.33
True Negative	1.71	2.02	1.60	1.60
False Positive	1.24	1.32	1.34	1.10
False Negative	1.34	1.74	1.38	1.26

Table 1: Mean x-flips for each sentence type in the original experiment and replication.

	No Context			Context		
	Estimate	Std. Err.	t value	Estimate	Std. Err.	t value
Intercept	1.21	0.12	10.45	1.09	0.13	8.62
Sentence Type (Negative)	0.36	0.19	1.87	0.05	0.13	0.42
Truth Value (True)	0.03	0.16	0.16	0.15	0.14	1.01
Sentence type x Truth Value	0.26	0.25	1.03	0.18	0.19	0.97

Table 2: Coefficients, standard errors, and t-values for the full maximal mixed effects models run on the x-flips data from the replication experiment.

Acceleration Components:

In Dale & Duran (2011) Experiment 1 (no context), a significant effect of sentence type was found on number of acceleration components, such that negative sentences produced a greater number of changes in acceleration than positive sentences. In addition, a significant interaction between sentence type and truth value was found, such that true positive sentences showed the fewest acceleration components but true negative sentence showed the most. In Experiment 3 (context), a smaller but still significant effect of sentence type was found, again with negative sentences producing more changes in acceleration, but the interaction between sentence type and truth value was no longer significant. The original means are plotted in Figure 3.

For the replication, acceleration components were originally calculated on the full set of x- and y- coordinates. However, this resulted in mean AC much higher than those reported in the original paper (see Table 3). After inspecting mouse trajectories and acceleration, it appears that most of these acceleration components occurred as the trajectory approached the decision point (i.e. the “true” or “false” box). Inspection of the original analysis code suggests that the original authors trimmed the data to a smaller window of x- and y- coordinates. If the window used for analysis did not include the “true” and “false” buttons, this could explain the differences seen in means between the original experiment and the replication. I trimmed the data to only include x- and y- coordinates such that y coordinates were more than 200 px, which eliminates the portion of the trajectory within 100px of the true/false buttons. This resulted in AC means closer to those reported in the original experiment; trimming the data in this way did not have an effect on the overall pattern of data. However, it is important to note that this was a somewhat arbitrary decision based only on a guess about what caused the discrepancy in means. A graph of both the trimmed and untrimmed data can be seen in Figure 3 (note the difference in the scale of the axes for the untrimmed data).

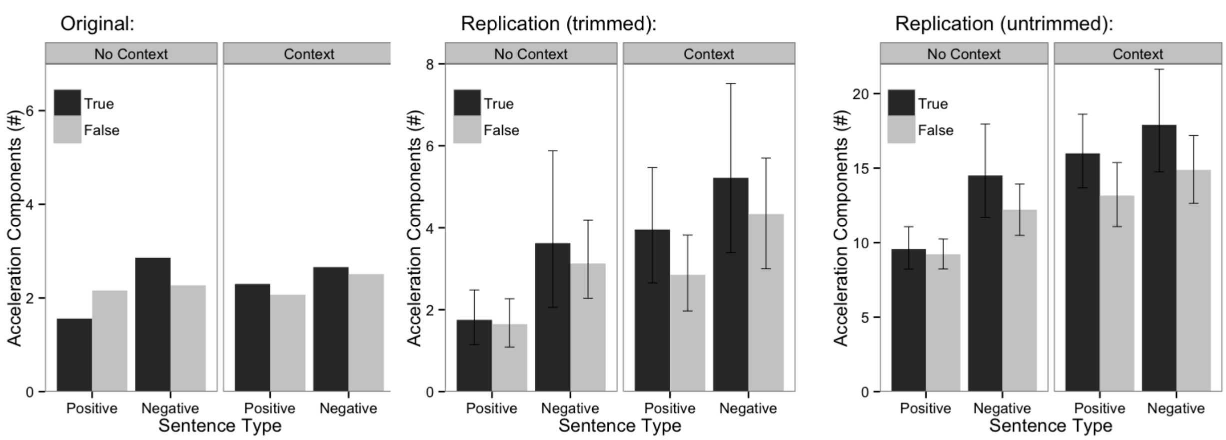

Figure 3: Original means (left) and replication means (right) for number of acceleration components. Error bars for replication data are bootstrapped 95% confidence intervals.

	No Context			Context		
	Estimate	Std. Err.	t value	Estimate	Std. Err.	t value
Intercept	8.66	0.59	14.67	12.07	0.88	13.77
Sentence Type (Negative)	2.13	0.84	2.55	1.43	1.10	1.30
Truth Value (True)	-0.05	0.80	-0.06	1.69	1.08	1.57
Sentence type x Truth Value	2.14	1.05	2.04	0.67	1.56	0.43

Table 3: Mean AC for each sentence type in the original experiment, trimmed replication data, and untrimmed replication data.

Mixed effects models with random effects of subject and item were tested separately for the no context and the context condition[2]. These models were run on both the trimmed and untrimmed data. In the no context condition, a significant effect of sentence type was found, with negative sentences producing more acceleration components than positive ones. However, there was no significant interaction between sentence type and truth value. The same model run on the untrimmed data found a significant main effect of negation as well as a significant interaction between sentence type and truth value. In the context condition, a model run on the trimmed data showed a marginal effect of sentence type, with negative sentences producing more AC than positive sentences, but not other significant main effects or interactions. The model run on the untrimmed data showed no significant main effects or interactions. Results of the models run on the untrimmed data can be seen in Table 4.

	No Context			Context		
	Estimate	Std. Err.	t value	Estimate	Std. Err.	t value
Intercept	1.48	0.30	4.93	2.36	0.48	4.95
Sentence Type (Negative)	1.08	0.52	2.04	1.41	0.73	1.93
Truth Value (True)	0.02	0.43	0.04	0.61	0.66	0.92
Sentence type x Truth Value	0.53	0.64	0.83	0.04	1.01	0.04

Table 4: Coefficients, standard errors, and t-values for the full maximal mixed effects models run on the untrimmed AC data from the replication experiment.

Exploratory analyses

Effect of Context Condition:

Although the original paper did make interpretations based on the differences seen between Experiment 1 (no context) and Experiment 3 (context), they did not formally test for any difference in x-flips or acceleration components between the two experiments. I tested two further maximal mixed-effects models, examining the addition of context condition as another variable in my models. The x-flips model showed a main effect of negation, but no significant

[2] Due to the maximal model failing to converge, random effects structure for both models was as follows: (sentence.type*truth.val | subid) + (sentence.type+truth.val | item)

effect of context and no significant interactions (Table 5). The AC model, run on the untrimmed AC data, showed main effects of negation as well as context. However, the context condition actually showed a significant *increase* in the number of acceleration components, which is counter to the predictions made by the original paper, which suggested that context should lead to less discrete mouse trajectories.

	Estimate	Std. Err.	t value
Intercept	1.21	0.11	10.76
Sentence Type (Negative)	0.37	0.15	2.39
Truth Value (True)	0.02	0.14	0.15
Context Condition (Context)	-0.13	0.17	-0.77
Sentence Type x Truth Value	0.25	0.20	1.23
Sentence Type x Context Condition	-0.31	0.23	-1.34
Truth Value x Context Condition	0.12	0.20	0.61
Sentence Type x Truth Value x Context Condition	-0.07	0.30	-0.25

Table 5: A maximal mixed-effects model showing effects of context on number of x-flips.

	Estimate	Std. Err.	t value
Intercept	8.65	0.67	12.827
Sentence Type (Negative)	2.21	0.98	2.269
Truth Value (True)	-0.02	0.90	-0.03
Context Condition (Context)	3.41	1.00	3.42
Sentence Type x Truth Value	1.96	1.33	1.48
Sentence Type x Context Condition	-0.75	1.43	-0.53
Truth Value x Context Condition	1.75	1.32	1.32
Sentence Type x Truth Value x Context Condition	-1.34	1.96	-0.69

Table 6: A maximal mixed-effects model showing effects of context on number of AC (untrimmed data).

Effect of mouse type:

It is possible that the type of "mouse" being used, e.g. a traditional computer mouse versus a trackpad, could influence the discreteness of the trajectories produced. I collected information from participants about the way they controlled their cursor. 80 participants reported using a mouse, 20 participants reported using a trackpad, and 2 participants selected "other" (these two participants were excluded from this analysis). I performed t-tests on the subject means for both x-flips and AC. I found a significant effect of mouse use on both measures. Mouse users showed significantly more x-flips than trackpad users, M_{mouse}= 1.70, $M_{trackpad}$ = 1.03, t(49) = 3.6, p<.001. Mouse users also showed more acceleration components than trackpad users, M_{mouse}=

4.10, $M_{trackpad}$ = 2.40, $t(49)$ = 2.20, $p<.05$.

Given this difference, I re-ran the confirmatory analyses described above to see if removing all non-mouse users from the data would change the results in any way. With non-mouse users removed, the full sample size was 80 participants, with 46 participants in the no context condition and 34 in the context condition. However, the data with this smaller sample looked nearly identical to the full sample data (see Figure 4). Furthermore, the maximal mixed effects models reported above were re-run for the no context condition, and no major changes to coefficients or t-values was found. Thus, removing trackpad users from the data does not appear to change the interpretation of the replication. For this reason, and due to the reduced sample size in the mouse-only data, the remaining additional analyses will use the full sample.

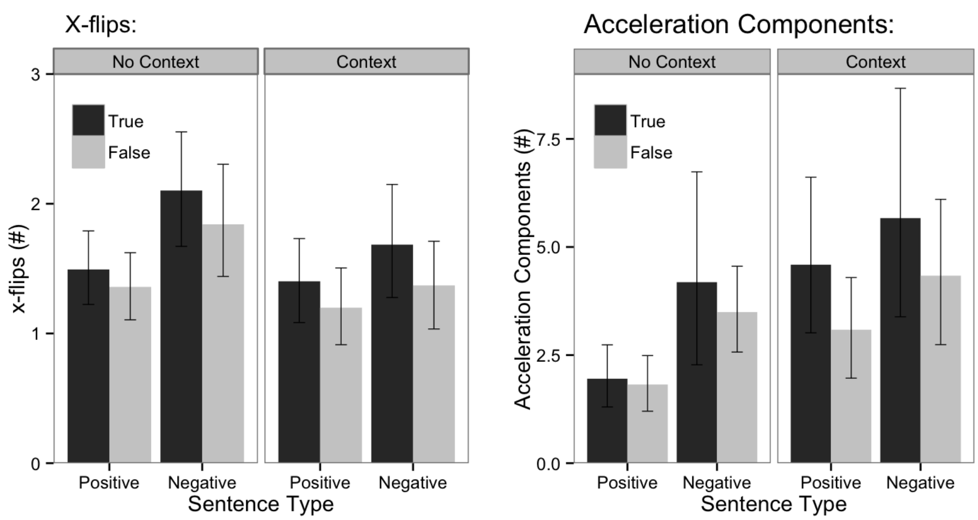

Figure 4: Graph of mean x-flips and acceleration components (trimmed data) for only participants who reported using a traditional mouse. The pattern of means is nearly identical to that seen in the full sample.

Midline Crosses:
In addition to testing the number of x-flips (the number of times the mouse changed direction along the x-axis), I analyzed a further measure, the number of times the mouse trajectory crossed the midline of the x-axis. This is a coarser measure of discreteness and captures the number of times participants completely changed the trajectory of the mouse. A graph of the number of these midline crosses can be seen in Figure 4.

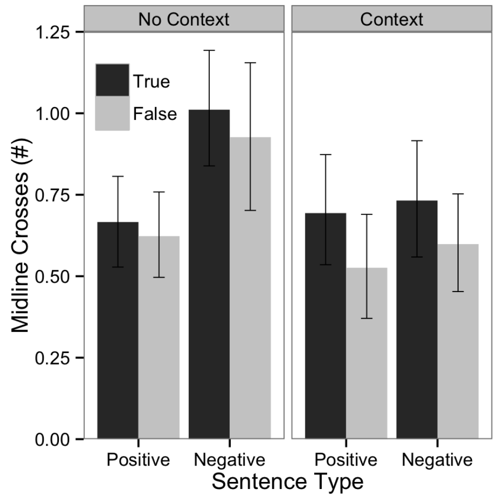

Figure 5: Number of times the mouse crossed the midline of the x-axis. Error bars are bootstrapped 95% C.I.

Mixed effects models were run to test the effects on sentence type, truth value, and their interaction on the number of midline crosses[3]. In the no context condition, a significant effect of sentence type was found, with negative sentences producing more midline crosses than positive sentences, but there was no main effect of truth value and no interaction between sentence type and truth value. In the context condition, there were no significant effects of either variable or the interaction on midline crosses. The results of this model can be seen in Table 7.

	No Context			Context		
	Estimate	Std. Err.	t value	Estimate	Std. Err.	t value
Intercept	0.57	0.07	8.42	0.53	0.07	7.10
Sentence Type (Negative)	0.25	0.10	2.59	-0.01	0.08	-0.07
Truth Value (True)	-0.002	0.10	-0.03	0.14	0.11	1.23
Sentence type x Truth Value	0.13	0.12	1.15	0.001	0.13	0.13

Table 7: Results of a maximal mixed effects model examining the effects of sentence type and truth value on the number of midline crosses.

Time to initiate mouse movement:

One potential problem with collecting mouse trajectory information on MTurk (or in the lab) is that you have little control over how participants initiate their mouse movement. If participants are pausing to process the sentence and make their decision before they initiate the mouse movement, it is possible that these participants are more likely to make smooth trajectories, with no AC or x-flips. I re-plotted the data, removing participants who took more than 2461 ms to

[3] For the no context model, the maximal model would not converge and the following random effects structure was used: (sentence.type+truth.val | subid) + (sentence.type+truth.val | item). For the context model, the maximal random effects structure was used.

initiate their first mouse movement, which was 2 standard deviations longer than the mean time to initiate first mouse movement. This resulted in slight changes to the overall pattern of data (see Figure 6).

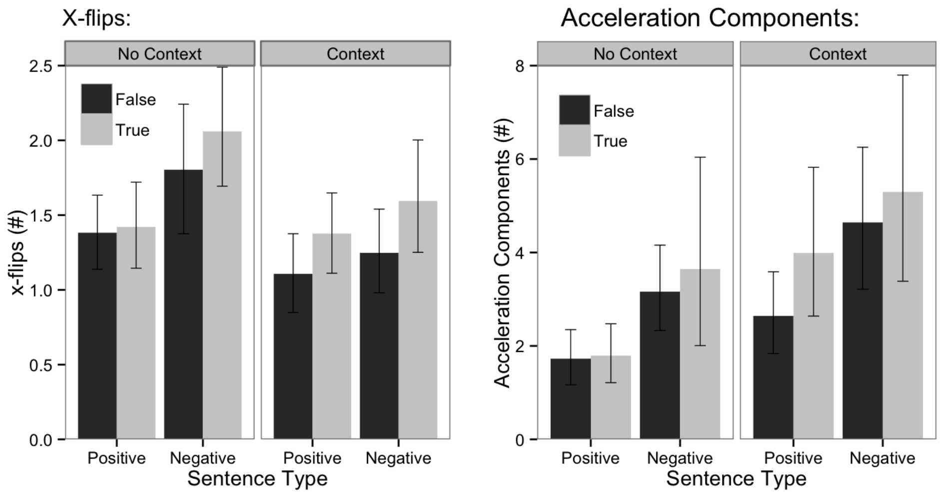

Figure 6: Mean x-flips and AC (trimmed data) for participants who began initiating their mouse movements quickly, i.e. ~2500ms after reading the final word.

In addition to removing participants who took longer to initiate their first mouse movement, I was also interested in whether sentence type, truth value, or context had an effect on how long it took participants to initiate their first mouse movement. Mean times to initiate mouse movement following presentation of the final word can be seen in Figure 7. A mixed-effects model was run to test the additive effects of sentence type, truth value, and context condition on time to initiate mouse movement.[4] There was a significant main effect of sentence type, with participants taking longer to initiate their first mouse movement after reading negative sentences compared to positive sentences. There was also a significant main effect of context, such that longer times to initiate mouse movement were seen following sentences in context compared to sentences in the no context condition. This is likely due to the fact that context sentences were longer, and may have led to longer processing times even after the final word was presented.

[4] The following random effects structure was used: (sentence.type+truth.val+context.condition| subid) + (sentence.type+truth.val| item)

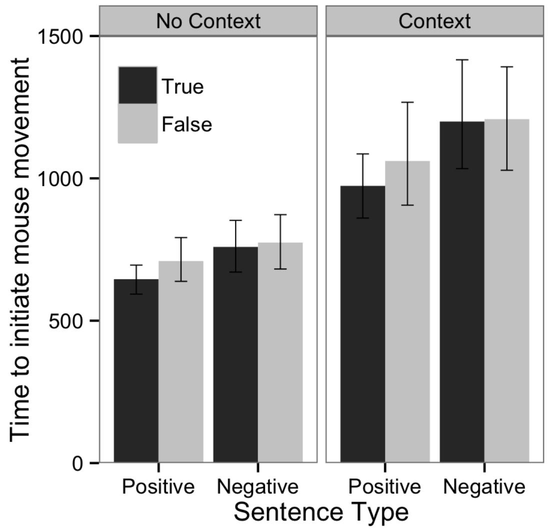

Figure 7: Mean time to initiate mouse movement following presentation of the final word.

	Estimate	Std. Err.	t value
Intercept	683	46	14.99
Sentence Type (Negative)	136	36	3.77
Truth Value (True)	-26	36	-0.73
Sentence type x Truth Value	306	65	4.8

Table 8: Results of a maximal mixed-effects model examining the effects of sentence type, truth value, and context on the time to initiate mouse movement.

Reaction Time:

Previous work on the processing of positive and negative sentences often uses reaction time as a measure (Wason, 1965; Clark & Chase, 1972; Glenberg, 1999). These previous studies have typically found similar interactions of sentence type and truth value (Clark & Chase, 1972), and similar effects of context (Wason, 1965; Glenberg, 1999). I compared the effects of sentence type, truth value, and context on reaction time, calculated as the time from the onset of the final word to the time that the participant selected true or false. Mean reaction times are plotted in Figure 8.

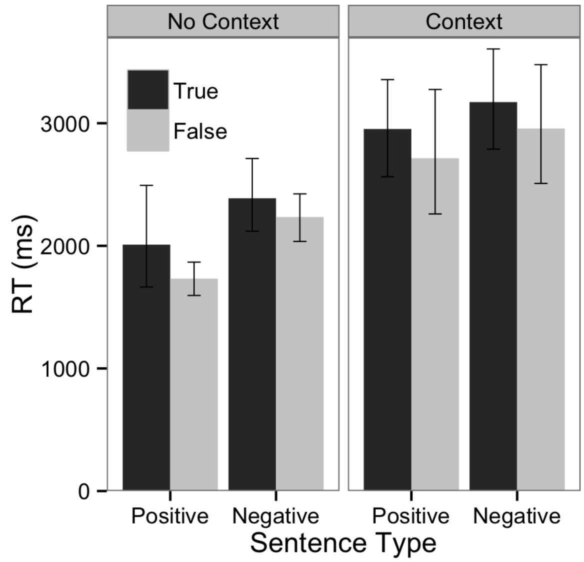

Figure 8: Mean reaction times (time from onset of final word to T/F selection)

I ran a maximal mixed effects model to test the additive effects of sentence type, truth value, and context on reaction time. There were significant main effects of sentence type, with negative sentences producing higher RTs, as well as context, with the context condition producing higher RTs (Table 9)**.** The effect of context is again likely due to the fact that context sentences were longer and may have led to longer processing times of the sentence as a whole (even if processing of the final portion of the sentence was facilitated).

	Estimate	Std. Err.	t value
Intercept	1653	125	13.19
Sentence Type (Negative)	344	98	3.51
Truth Value (True)	138	87	1.59
Context Condition (Context)	684	167	4.10

Table 9: Results of a maximal mixed effects model examining the effects of sentence type, truth value, and context on reaction time.

It is possible that these effects of negation and context can be explained by the effects of negation and context seen on time to initiate first mouse movement (Table 8), since RT includes the time it took to initiate the mouse movement. Looking at the beta values, it appears that delay in time to initiate mouse movement explains some but not all of this effect. For the effect of negation, the model of time to initiate mouse movement had an estimated effect of 136 ms, while the model of RT shows an estimated effect of 344 ms. For the effect of context, the model of time to initiate mouse movement shows an estimated effect of 306 ms, while the model of RT shows an estimated effect of 684 ms. Thus, it appears that some of the effect of negation and context occurs in the first few seconds, while participants are still processing the sentence and planning their mouse movement, but negation and context may continue to have detrimental effects on RT even after the mouse trajectory has been initiated.

Reading Time:
In addition to analyzing reaction times, I also collected information about reading times - i.e. the time each participant spent on each word in the sentence. From this I analyzed how long it took participants to read the entire sentence, except for the final word (since it is impossible to know when participants stopped reading the final word and started planning their mouse movement). A maximal mixed-effects model testing the effect of sentence type on total reading times was run for both the no context condition as well as the context condition. I did not look for any effects of truth value, since participants would not know the truth value of the sentence until they had read the final word. In the no context condition, a significant effect of sentence type was seen such that negative sentences produced an estimated increase in reading times of 593 ms ($t=3.7$). However, since the mean time spent reading negative words was 525 ms, this is likely due to the fact that negative sentences are slightly, but reliably, longer than positive sentences. In the context condition, negative sentences produced an estimated increase in reading times of 453 ms; however, this effect was not significant ($t=.56$), likely due to the fact that context sentences were much longer than no context sentences.

I also analyzed how long it took participants to read negative words (e.g. "no" or "not") in no context vs. context. There was no significant difference between reading time of negative words in the no context condition and the context condition, $M_{no\ context}$ = 531 ms, $M_{context}$=511 ms, $t(100) = .68$, $p=.5$. A more interesting analysis would be to compare reading times of negative words to other words; however, this would require designing a different experiment which controlled for the length of other, non-negative words.

Discussion

Summary of Replication Attempt

Dale & Duran (2011) examined mouse trajectories in response to true or false, positive and negative sentences, either without any contextual support (Experiment 1, no context) or with the added context of embedding the sentence within a quote spoken from an adult to a child (Experiment 3, context). They found that when sentences were presented without context, mouse trajectories were more discrete (based on the measures of x-flips and AC) in response to negative sentences, and an interaction was found between sentence type and truth value such that sentences were the smoothest in response to true positive sentences, and most discrete in response to true negative sentences. However, when the sentences were presented within context, the interaction between truth value and sentence type disappeared for both measures, and the main effect of negative sentences disappeared for x-flips and decreased (but was still present) for AC. In my replication, I replicated some of these results, but there were also important differences between my replication data and the original study

The measure of x-flips produced a marginal effect of negation, but the interaction seen in the original paper was not significant. Visually, the replication data look very similar to the original data, but there is too much variance to produce significant results. A power analysis by simulation, based on the effect estimates and variance reported in the mixed-effects model of the no context condition, showed that the replication attempt had power=0.55, suggesting that this study was underpowered. Further replication attempts could use larger sample sizes.

Our models of acceleration components showed very similar results to the original data, with a significant effect of negation and a significant interaction in the no context condition, but no significant main effects or interactions in the context condition. However, there are some problems with the replication data that make this difficult to interpret. The “untrimmed” data that produced these model results had much higher AC means than those seen in the original paper. Trimming the data to a smaller window produced more similar means, but this was a somewhat arbitrary decision and it is not clear whether the original authors trimmed the data before calculating acceleration components. Furthermore, in both the trimmed and untrimmed versions of the replication data, context condition produced a *higher* number of acceleration components. This was not the case in the original data, and it is not consistent with the authors’ interpretation of the original results, which is that embedding the sentences in context should reduce the discreteness induced by negative sentences.

Commentary

In addition to the replication analysis described above, I also conducted a number of additional analyses. Some of these additional analyses had interesting implications for the interpretation of this replication as well as considerations for future research. The most interesting and relevant of these findings are summarized below, with a discussion of potential implications.

Models of the data that included context as a factor (as opposed to analyzing the two context conditions separately, as was done in the original paper), did not show the expected effects of context. According to the original paper, the presence of context should lead to less discrete mouse movements in response to negative sentences; thus, these models should have shown either a main effect of context or a context x negation interaction. In the x-flips model, neither of these were significant. In the model of acceleration components, there was a significant effect of context but in the *opposite* of the expected direction, with the context condition showing significantly higher numbers of acceleration components. Given the fact that the replication data had very different means for AC than the original paper, it is possible that there is a discrepancy between how AC was calculated in the original and in this replication.

The original paper did not describe any requirements of “mouse type” (e.g. whether a traditional mouse or a trackpad was used). For this replication, I collected this information and found a significant effect of mouse type, with mouse users showing more discrete trajectories than trackpad users. However, removing trackpad users did not ultimately have an effect on the results. Still, it could be useful for future mouse-tracking studies to limit participants to those who use a traditional mouse.

I also completed additional analyses of time to initiate mouse movement (i.e. time from onset of final word to first mouse movement) as well as overall reaction time (i.e. time from onset of final word to selection of true or false). Significant effects of negation and context were seen in both of these measures. Although it is possible that the effects on reaction time are due entirely to latency to initiate mouse movement, this did not appear to be the case; it appears that negation and context led to longer reaction times independent from the increase in time to initiate mouse movement. The significant effect of negation on these measures is theoretically interesting; the fact that negative sentences incur greater processing times has been found in a number of previous studies (e.g. Clark & Chase, 1972). The effect of context on these measures, however, was surprising; previous studies of the effects of context on processing these types of sentences suggests that context facilitates processing, leading to a *decrease* in RT (e.g. Wason, 1965; Nordmeyer & Frank, in prep). It is possible that the self-paced reading procedure did not give participants enough time to process the sentences, thus leading to differences in RT purely due to the length of the sentences (since the context sentences were much longer). If this is the case, a procedure which uses a visual context (e.g. Nordmeyer & Frank, in prep) rather than a verbal context could be preferable, since the time it takes to process the context is more controlled.

Bibliography

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

Dale, R. & Duran, N. (2011). The cognitive dynamics of negated sentence verification. Cognitive Science, 1-13.

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

Wason, P. (1965). The contexts of plausible denial. Journal of Verbal Learning & Verbal Behavior, 4(1), 7-11.