i. Introduction

This experiment was a replication of an experiment out of the University of Glasgow. The primary goal of this study was to investigate the level of abstraction involved in the mental structures we use to represent sentences during language processing. Specifically this study and my replication sought to determine whether the structures of mathematical equations and sentences may be stored in overlapping ways suggesting a domain general level of abstraction in human language processing and mathematical processing.

These questions were investigated using the observation that certain sentences and certain equations have similar structures. Specifically sentences of the form “noun phrase 1 did some verb to noun phrase 2 of noun phrase 3 who/that _____.” For example “I visited a friend of a colleague who lived in Spain.” Importantly these sentences are ambiguous between two interpretations. The relative clause, “who lived in Spain” can be low attachment, applying to the colleague or it can be high attachment applying to the friend of a colleague. There are also equations that mirror this structure in a way. We can take two equations of the form $A\ +\ |\ -\ (\ (\ B\ +\ |\ -\ \ C\ )\ \times\ |/\ D)\ =$, $A\ +\ |\ -\ \ B\ +\ |\ -\ \ (C\ \ \times\ |/\ D)\ =$. For example $80\ -\ (\ (\ 9\ +\ \ 1\ )\ \times\ 5)\ =$ and $80\ -\ 9\ +\ (1\ \times\ 5)\ =$. The first type mirrors the high attachment sentence because the times five applies to the entire nine plus one just as the relative clause can attach to the entire “friend of a colleague.” Meanwhile the second equation mirrors the low attachment sentence interpretation because the times five applies to just the one just as the relative clause could apply to just the colleague. We sought to investigate the domain generality of abstract syntactic representations by priming experimental

participants with equations of those to forms to see if that affected their interpretation of relative clause attachment in ambiguous sentences.

ii. The Experiment

Methods

Participants

Using prolific I hired 50 participants with the following criteria. I asked that their country of birth and country of current residence was the United States and that their first language was English. In addition I barred participants who had Prolific approval ratings below 95%. Further information was gathered about the participants using a short survey at the end of the experiment. All fifty participants filled out the survey. Their self-reported ages ranged from 19 to 66 years old, with the median age being 29.5 years old. 24 reported their gender as male, 24 female, and 2 as other. They also represented a range of different education backgrounds and levels. Each was paid 5$ for an expected task time of 20 minutes.

Materials

12 experimental relative clause fragments were created for this experiment. Importantly these sentences had the relative clause entirely removed, leaving just a “who” or “that” and a blank. For example “The secret service confiscated all files of the organisation that ____ .” One further note is that one of two noun phrases that the relative clause could attach to was always plural and one was always singular. In addition either both noun phrases were animate people or both were inanimate objects.

In this way it was syntactically ambiguous which of the two the relative clause would attach to.

12 triplets of three mathematical expressions were also created. One of the three was a “high attachment” equation taking the form $A\ +\ |\ -\ (\ (\ B\ +\ |\ -\ \ C\ )\ \times\ |/\ D)\ =$. The second was a “low attachment” version that was nearly identical except that the double parentheses were switched out for a single set around the $C\ \times\ |/\ D$ part. The third equation was baseline equation that took the form $E\ +\ |\ -\ |\ \times\ |/\ F\ =$. For example, one mathematical triplet was the following:

$3\ +\ (\ (6\ -\ 2\ )\ /\ 2\ )\ =$

$3\ +\ 6\ -\ (\ 2\ /\ 2\ )\ =$

$12\ /\ 6\ =$

For each participant one equation of each of the triplets was randomly selected, such that each participant was assigned 4 high attachment equations, 4 low attachment, 4 baseline. Then each of these 12 equations, selected for each of the participants, was assigned to a random one of the sentence fragments described earlier, to create 12 experimental sentence fragment-equation pairs.

In addition there were 24 filler equations and 24 filler sentences fragments. These filler equations and sentence fragments took a variety of different forms. Importantly, each of the 24 sentence fragments consisted of just the first part of a sentence followed by a blank and none of them ended in a who/that relative clause like those from the experimental set. Additionally, like the experimental equations, all filler equations consisted of only addition, subtraction, multiplication, and division and all ended in an equals sign with no answer on the right side of the equals sign. For each

participant the 24 filler equations were randomly assigned to the 24 filler sentences to create 24 filler sentence fragment-equation pairs.

These 12 experimental sentence pairs and 24 filler pairs created 36 randomly created pairs unique to each participant. These 36 pairs were then shuffled into a random order for each participant.

Procedure

Participants were first prompted with an instruction screen. On the instruction screen it was described to them that they would be alternating between two subtasks. The first subtask was to solve equations when prompted and the second was to provide completions to sentences that had been “cut short.” Upon reading this and clicking a start experiment button the task would begin.

As described in the materials section, a unique randomly ordered set of 36 pairs of equations and sentence fragments was created for each participant. As they went through the task section they would be prompted with the equation and then a sentence fragment from each pair. When prompted with an equation the slide on screen would read “Please provide the answer to the following equation.” Followed by the equation, a textbox to enter their answer, and a continue button. They could only continue to the next slide upon entering an answer. If they tried to continue with the textbox blank the slide would read “Make sure you have entered an answer before hitting the 'Continue' button.” The experiment would continue regardless of whether their answer was correct or not. When prompted with a sentence fragment the screen would read “Please fill in the blank to complete the sentence.” This was followed by the sentence fragments which always ended in a blank line, “_____”, followed by a period. Beneath this was

again a textbox to enter their completion and a continue button. As with the equations they could not continue to the next slide without entering something into the textbox. At no part in the experiment could the participants go back to change answers.

After completing their 36 pairs of equations and sentence completions they were prompted with an optional after-experiment survey. Participants were asked if they believed they did the experiment correctly, whether they thought the payment was fair, and whether they enjoyed the experiment. They were also asked the following demographic data: gender, age, level of education, and native language.

After all participants finished the experiment I went through each of their sentence completions for the 12 experimental sentence fragments. I then flagged each sentence completion as high attachment or low attachment based on which noun phrase the participant had created a relative clause for. In order to determine which noun phrase the completions attached to I first looked at cases of marked number. Since every sentence fragment I provided had one plural and one singular marked number in the sentence completion could determine whether it was high or low attachment. For example if the sentence fragment “The minister saw the bodyguard of the diplomats who____” was completed “were visiting the town, plural number of “were” tells us that this attaches to diplomats rather than bodyguard and is therefore low attachment. If the number was not clear from the completion I used broader contextual clues. For example the sentence “The tourist guide mentioned the bells of the church that ____” might be completed with “chimed every day.” Though “chimed” does not mark number it clearly applies to the bells and is therefore high attachment. Those that were not determinable from grammatical or contextual clues were marked as unclassifiable.

Using these markers of attachment I created 12 datapoints per participant. Where a datapoint consists of A) the priming condition, i.e which type of equation preceded the target sentence, baseline, high attachment, or low attachment, and B) sentence completion type, high attachment, low attachment, or unclassifiable. However, those data points with unclassifiable completions were excluded.

In addition those trials for which the participant incorrectly solved the priming equation were also excluded. This is because if the participant gets the answer wrong this may suggest that they misread the equation or otherwise misrepresented its structure in their mind. This would invalidate it as a prime for the target sentence.

Deviations from original experiment

There were numerous differences between this experiment and the original version. I will detail the largest and most important. The first difference was in the participant sample. The original experimental participants were University of Glasgow psychology students, whereas my participants were American and represented a broad range of ages and educational levels.

The second key difference is that my experiment was conducted online using Prolific whereas the original experiment was conducted in person using a pen and a paper booklet. This difference is very important because in the original experiment they were able to prevent participants from using calculators altogether. I had no way to fully prevent calculator usage in my experiment. I did ask the participants in the initial instructions tab not to use a calculator. In addition I used photos of the equations rather than typed equations so that it was not convenient to copy paste the equations into a calculator in another computer window. To further disincentivize calculator usage I made

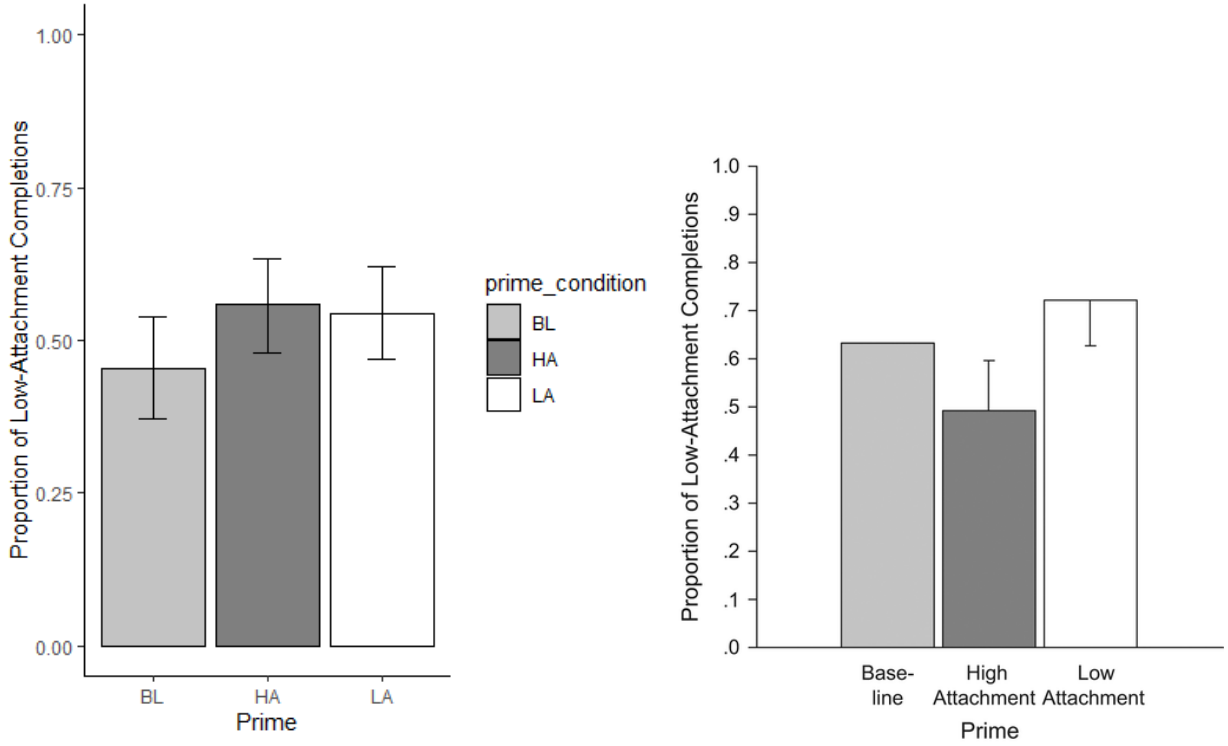
the equations easier than in the original. All equations were borrowed from the original experiment, but the numbers were lowered in order to make mental arithmetic easier.

Another key difference was that I reduced the number of target sentence, priming equation pairs from the 24 of the original down to just 12. There were two reasons for this. A) Some of the 24 sentences seemed more likely than others to generate unclassifiables, which creates lots of unusable data B) by shortening the experiment participant fatigue would be reduced, hopefully further reducing desire to use a calculator. I contacted a researcher from the original study and he was able to tell me several sentences that lead to particularly high rates of unclassifiables. I also removed several more that I personally expected to generate unclassifiables. An unintended consequence of this is that there was half as much data collected per participant. In order to account for this halving of data I increased the participant number from 27 in the original to 50 in this version.

Results

With 50 participants and 12 experimental pairs each there were initially 600 data points. After excluding trials with unclassifiable sentences as well as trials with incorrectly solved priming equations this number was reduced to 451 data points. Of these 31.7% were baseline priming equations, 33.3% were high attachment primes, and 35.0% were low attachment primes. 47.9% of target sentences were completed with a high attachment completion and 52.1% with a low attachment. As summarized, in the below chart 45.5% of sentences were completed with a low attachment clause when the prime was baseline, 60.4% when the prime was high attachment and 58.9% when the

prime was low attachment. This is shown on the left alongside the results of the original experiment on the right. As can be seen the results are ordered in a very different way. In the original high attachment primes correlated with fewer low attachment completions than the baseline, and low attachment primes with more than the baseline. In my experiment low attachment primes also correlated with more low attachment completions than the baseline, but the high attachment primes correlated with even more.



I also ran a mixed effects logistic regression model. The only fixed effect was the dependent variable, the prime condition. There were two random effects in the model, specific participant as a slope term, and specific target sentence fragment as an intercept term. The result of this mixed effects model was that the p values for high attachment and low attachment primes affecting the completions were .25 and .26 respectively. Both were significantly higher than the .05 threshold. This is a

nonreplication of the original as neither condition showed a statistically significant effect and what statistically insignificant effect there was followed a different pattern to the original as well.

iii. Discussion

In summary, this experiment sought to investigate whether the structures of equations could prime participants' interpretations of the structures of similar sentences. In doing so it looked to determine if there was a domain general level of abstraction at which both sentence structures and equation structures are stored. No statistically significant effect was discovered and so no evidence for such a level abstraction was found. This is a non-replication of the original experiment which did find such an effect.

There are several possible explanations for why this experiment failed to replicate the original. One is that the original experiment was just a fluke result. The original experiment did not use a mixed effects model to analyze its data and it is possible that doing so would have resulted in insignificant p values as well. Alternatively it could be a product of the modifications made to the experiment. Perhaps the most significant and likely suspect would be the change to an online format, which made it far harder to monitor and prevent calculator usage. It would be ideal to see this experiment replicated in a lab setting more similar to the original.

One possible modification that could be made to this experiment in a future replication would be to change the target sentences from the past tense to the present. The past tense tends to lead to more forms that are unmarked for number. Since

grammatical number was used to classify the attachments this resulted in a higher rate of unclassifiables than might be expected with present tense sentences.

iv. Original experiment citation:

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural Priming Across Cognitive Domains: From Simple Arithmetic to Relative-Clause Attachment. Psychological Science, 22(10), 1319–1326. https://doi.org/10.1177/0956797611416997