Teacher Note:
Very nice study! Well conducted and clearly written. Nice with tables and graphs! Judging by the accuracy scores, it seems that you experienced something like a floor effect - that the task for simply too difficult (which is also interesting).

Research Report

Logical Reasoning Experiment

Ida, Sebastian, Signe KK, Majka, Lukas & Signe H
Cognitive Science, Aarhus University

Abstract

This paper examines whether the content of syllogisms and conditionals affects people's reaction time in evaluating their logical validity. We used conditional and categorical syllogisms of the exact same logical structure, but with differences in content, resulting in 3 conditions: abstract/no content, meaningless content, and meaningful content. We carried out a between-participants design, where each participant had to answer 10 syllogisms in one of the conditions. Their reaction time was measured for each syllogism, and analysis was carried out on all trials with correct answers with the mean reaction time for each participant.

There was found no significant difference in reaction time for the three different conditions, suggesting that the content of syllogisms has little to no effect on people's judgment time.

Keywords: Reasoning; Syllogisms; Logic; Meaning; Conditionals

Introduction

We humans have used our brains to make endeavours into the realms of logical thinking for thousands of years. One of the earliest examples hereof the famous syllogism, formulated by Sextus Empiricus roughly 2000 years ago:

"Everything human is an animal.
But Socrates is human.
Therefore, Socrates is an animal"
(Annas & Barnes, Outlines of Scepticism, 2000, II.:120).

Socrates is since long dead and gone, and most people today are capable of following the logic in this example. Still, it is often claimed that humans aren't inherently logical beings, and that our cognitive abilities are more than lacking in this field (Anderson, 2015). This is especially evident when it comes to the kind of strict, abstract logic required in syllogisms that involves complicated, deductive reasoning - a discipline in which computers and AI are far superior to us. Why is this? Are some logical statements easier to judge than others?

A syllogism is an argument "consisting of two premises and a conclusion" (Anderson, 2015: 238). The earlier example is an instance of a categorical syllogism.

A conditional syllogism takes the form of an if-then-statement; IF a condition is met (the antecedent), THEN a consequence follows (the consequent) (Anderson, 2015). The third part is a conclusion which must be judged in accordance to these two premises (see example in Table 3).

Relevant literature

There are several factors that may cause people's thinking to deviate from traditional logic.

It has been shown that the structure of a logical statement has a significant effect on people's accuracy in judging its validity, with some logical constructions being less intuitive

and harder to judge for most people (Evans, 1993).
Additionally, it has also been discovered that people are better at judging the validity of a logical statement in a real world scenario compared to an abstract one (Evans, Handley, & Harper, 2001). People tend to accept any real world syllogism with a believable conclusion. Only if the conclusion denies their belief, they check on the logic. In relation to this, Johnson-Laird proposed a mental model theory which states that while judging the possibility of the conclusion, participants create a mental model of the world that satisfies the premises before examining that model to see if the conclusion is satisfied (Johnson-Laird & Steedman, 1978). However, they are not considering all possible models that would satisfy the premises. This can be considered as another example of why people often fail in reasoning.
Furthermore, it has been discovered that the content of logical statements (which is completely logically irrelevant) influences which brain areas are activated (Goel, Buchel, Frith & Dolan, 2000). Parietal regions are more active when participants are solving content-free material, and left prefrontal and temporal-parietal regions associated with language processing are active when judging meaningful content.

Research questions & hypotheses
Building on these ideas that different types of content in syllogisms fosters different kinds of brain activation and thinking strategies, we will focus on differences in processing speed. We hypothesize that syllogisms with meaningful content, where strict, unintuitive logical reasoning is not necessarily needed are solved quicker by participants than syllogisms with strictly abstract content. We also hypothesize that syllogisms with 'meaningless' content (where real world logic and intuition doesn't apply) are solved slower than the meaningful syllogisms, but still quicker than the abstract syllogisms. An argument supporting this hypothesis is that the content of the non-abstract syllogisms (meaningful or not) may foster mental imagery in the minds of people, making the syllogism more tangible and good for creating mental models of different scenarios, thus enabling them to reach a conclusion faster.
In short, our hypothesis for reaction time in the different conditions is as follows:
Abstract (1) > Meaningless (2) > Meaningful (3), with reaction time being slowest in the abstract condition (1).

Materials and Methods

Participants

Partici-pants	Condi-tion 1	Condi-tion 2	Condi-tion 3	Total
Female	5	8	8	21
Male	6	3	2	11
Total	11	11	10	32

Table 1: Participant distribution across conditions

The age of the participants varied from nineteen to thirty-two years of age, with means and standard deviations for each condition as follows:

	Condi-tion 1	Condi-tion 2	Condi-tion 3
Mean age	24	22	23
Stan-dard devia-tion	3.86	2.11	2.41

Table 2: Means age (rounded to whole numbers) and standard deviations for participants

Participants with prior knowledge of the experiment and of logical statements in general, such as other Cognitive Science students, were excluded from the experiment.

Materials/Stimuli

As discussed earlier, the experiment consisted of three conditions: an abstract condition, a meaningless condition, and a meaningful condition. Each participant went through only one of the conditions (a between-participants design). 10 logical statements, both conditional and categorical syllogisms (with the exact same logical structure across conditions) were presented to the participant.

Condition 1 (Abstract)	Condition 2 (Meaningless)	Condition 3 (Meaningful)
If A, then B. A does not apply. → B doesn't apply.	If you have red hair, you can't swim. You do not have red hair. → You can swim.	If you drink poison, you die. You do not drink poison. → You do not die.

Table 3: Examples of the same conditional syllogism across all 3 conditions *(In this case, the statements are FALSE, since the consequent does not follow directly from the premises)*

Each participant's reaction time for judging whether a statement was true or false was measured for each of the 10 statements.

Procedure

We used PsychoPy2 to create our experiment, presenting the stimuli to our participants and collecting their reaction time and accuracy for each statement, including individual participant information such as name, age etc. Stimuli were presented in a fullscreen window with a white background, and the text was printed in black.

Participants first had to fill out a pop up window with their individual information before starting the experiment. They were then presented with a short intro text with useful instructions. The participant pressed a random key to start the experiment. Then the participant was engaged in a session consisting of 10 statements in the specific condition. They used the keyboard to press either A (to accept) or D (to deny) each statement, reflecting whether they thought the conclusion followed logically from the two premises or not.

Analysis

We used R Studio to carry out the analysis. The raw data consisted of ten data points for each participant (the reaction time for each statement). First we filtered out all the incorrect answers, only focusing on the correct answers. Since we are doing a one-way ANOVA and not a mixed-model analysis, each data point needs to be independent to meet the assumptions of the model. Therefore we calculated the mean reaction time for each participant from their correct trials. This way we ended up with a single independent data point (mean RT) for each participant.

First, we went through the assumptions check for the one-way ANOVA and found non-linearity and non-normality in the residuals. This led us to carry out a reciprocal transformation on our data. The transformation made the assumptions of the ANOVA tenable.

We conducted an independent, one-way ANOVA with condition as predictor and reaction time as outcome variable. The abstract condition was set as the baseline level and in this way we compared the two other conditions against this baseline. We also carried out a pairwise post-hoc test to compare each condition mean against the others and see whether there was a significant difference between any of the conditions. We used the conservative Bonferroni post hoc test in order to minimise the risk of type 1 errors.

Results

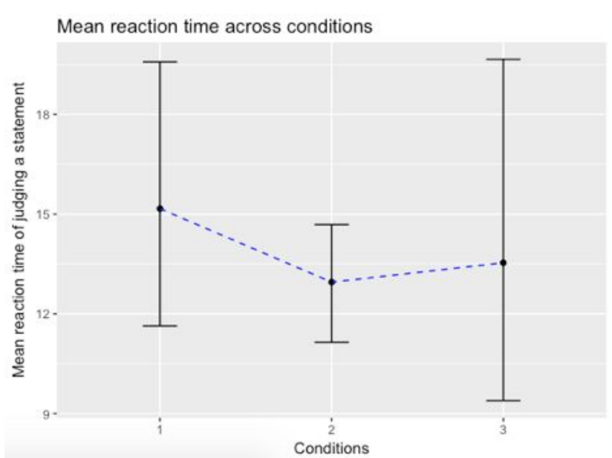

Figure 1: Mean reaction time in seconds (non-transformed) by condition (15.16 (sd = 7.05), 12.95 (sd = 3.26), 13.53 (sd = 8.65))

	Condi-tion 1	**Condi-tion 2**	**Condi-tion 3**
Mean reciprocal reaction time	0.078	0.082	0.092
Stan-dard devia-tion	0.031	0.022	0.034

Table 4: Mean reciprocal reaction time and standard deviations for each condition

There was found no significant effect of conditions of meaning on reaction time. $F(2, 29) = 0.60$, $p = 0.56$

This means that there was no significant difference in reaction time between the three conditions.

The Bonferroni post hoc tests also revealed that there was no significant difference in reading time in each condition compared to the other conditions. All p-values were $p > .05$.

Conditions	1	2
2	1.0	
3	0.87	1.0

Table 5: Exact p-values between conditions in the post hoc test

Discussion

The experiment showed no significant difference in reaction time between conditions. Therefore, our hypothesis - that statements with meaningful content are solved faster by participants than statements with abstract content - cannot be confirmed by our findings.

Methodological limitations
We used an independent ANOVA test to see if there was a difference between the means of reaction time in our 3 conditions. By using a means model we're forced to throw away a lot of data points (from 211 datapoints to 32) and thereby also a lot of within-participant variance. Using a mixed effects model would solve this problem by letting us designate some of the within-participant variance as specific random effects and thereby account for it, possibly reducing the size of the error bars.
An example of a random effect we could have accounted for is the fact that people have different base levels. Some of the participants may be more experienced in solving logical tasks or show difference in patience with this kind of task.

The amount of participants also affected our results. We have large confidence intervals that overlap between conditions, which might have been reduced if we had more participants.

Further Research
We've only looked at the reaction time of the participants, but it could also be interesting to find out more about their accuracy. Looking at the structure of the logical statements where most people fail might give an insight into where more research could be done:

D_mfull.png = 1 out of 10 answered correctly
A_mfull1.png = 1 out of 10 answered correctly
D_mfull5.png = 3 out of 10 answered correctly
A_mless2.png = 0 out of 11 answered correctly
D_mless5.png = 4 out of 11 answered correctly
D_abs5.png = 2 out of 11 answered correctly
D_abs9.png = 4 out of 11 answered correctly
(D: deny, A: accept, mfull: meaningful, mless: meaningless, abs: abstract)

The only case where the majority of people answered incorrectly across all conditions was with the 5th type of syllogism:

"Some x's are y's
Some y's are z'
Some x's are z's"

To test whether people were actually significantly worse at this syllogism, more trials could be arranged for the experiment. Or the experiment could be altered with new text-examples than the one used in this experiment:

"Some astronauts are female
Some females are gingers
Some astronauts are gingers"

to see if the misconceptions were solely due to the choice of words.

We did not have access to brain scanning equipment, which could give insight into whether or not the reaction times of our conditions have a significant relationship with activation of different brain areas, as noted in the introduction (Goel, Buchel, Frith & Dolan, 2000). It could be interesting to compare the activation of these areas with our conditions to see if the left prefrontal and temporal-parietal

regions associated with language processing were active when judging meaningful content. In addition to this we could also switch from looking at reaction times to looking at accuracy, conducting the experiment in a way where each syllogism was shown for a set time of e.g. 60 secs before the participant was allowed to answer to see if this would force people to reevaluate their mental models.
All in all, it seems like many different factors influence humans' speed and accuracy when forced into logical reasoning.

References

Anderson, J. R. (2015). *Cognitive Psychology and its implications*, 8, 237-259

Annas, J & Barnes, J, Empiricus, S. : Outlines of scepticism (2000)

Evans, J. S. B. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision.

Evans, J. St. B. T., Handley, S. J., & Harper, C. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. Quarterly Journal of Experimental Psychology, 54A, 935–958.

Goel, V., Buchel, C., Frith, C., & Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage, 12, 504–514.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. Cognitive Psychology, 64–99