No Effect of Semantic Congruency in a Masked Congruency Paradigm

Author:	Jennifer March (s3218163)
Email:	j.march@student.rug.nl
Project:	Minor Thesis
Degree Programme:	MSc Behavioural and Cognitive Neuroscience (C-Track)
Faculty:	Science and Engineering
Institution:	University of Groningen
First Supervisor:	Professor Sebastiaan Mathôt
	Behavioural and Social Science, Experimental Psychology
Second Supervisor	Professor Karin Pilz
	Behavioural and Social Science, Experimental Psychology
Date:	22.11.2020

Abstract

Embodiment theories of language suggest that parts of our perception and action systems are involved in understanding language. While neuroimaging and eye tracking studies provide evidence that parts of our perception and action systems are involved in understanding single words, the processing of more complex phrases is less well understood. The congruency effect is a slowing of responses for incongruent trials compared to control and congruent trials. Similarly, negation effect is a slowing of responses for negated phrases compared to their affirmative counterparts. Both congruency and negation studies claim ‘conflict’ to be responsible for this slowing. In the present study we further investigated the nature of this conflict in a masked semantic congruency paradigm. In our reaction time experiment, we asked participants to distinguish between animal related words and other words. The experiment consisted of two blocks one with and one without negations. In the latter, semantic congruency was realized by presenting words related to brightness in either white (i.e. congruent) or black (i.e. incongruent) font, while the reverse applied to negated block. For example, the word ‘sun’ presented in white was defined as congruent and the phrase ‘no sun’ in white font was defined as incongruent. We predicted that in negated incongruent conditions two types of conflict would cancel each other out leading to a decrease in reaction time. Our results indicate that there was no difference in reaction time for congruent and incongruent trials. While we found slower reaction times in the negated block, this did not depend on congruency. We propose that the conflict arising from negations should be distinguished from that arising of incongruency.

Keywords: Embodied Cognition, Stroop-task, Negations, Conflict

No Effect of Semantic Congruency in a Masked Congruency Paradigm

Language is a key part of being human. In comparison to animal communication, human language goes beyond identifying members of a species and designating items of immediate biological significance. It facilitates the formation and articulation of ideas and enables cultural transmission (Cancho & Solé, 2001). Recent neuroimaging studies suggest that language is represented by a various brain regions, beyond linguistic areas (e.g. Ben Shalom & Poeppel, 2008; Pulvermüller, 2013; Willems & Casasanto, 2011). This is in line with, embodiment theories of language, which suggest that parts of our perception and action systems are involved in understanding language (Willems & Casasanto, 2011). Specifically, to understand the meaning of words, brain regions associated with a word's functionality are activated (Galetzka, 2017). For example, the word 'kick' triggers a similar brain response to its action Similarly, visual perception studies have demonstrated that words semantically related to darkness (e.g. 'night') or brightness (e.g. 'sun') evoked pupil dilation or constriction respectively (Mathôt, Grainger & Strijkers, 2017, Mathôt, Sundermann & van Rijn, 2019).

The effect of altering the meaning of a word by syntax is less well understood (Meteyard, Rodriguez Cuadrado, Bahrimi & Vigliocco, 2012). Yet, this may be more apt to real live situations in which people are confronted with strings of words, or sentences, rather than single words. In the present paper, we aim at establishing behavioural evidence that syntax is embodied. Specifically, we examine the effect of negations in a semantic congruency paradigm.

Congruency Paradigms and Embodiment

A common and highly influential congruency paradigm is the Stroop task (Stroop, 1935). In Stroop's (1935) best-known experiment, participants were required to identify the print colour of words as quickly and accurately as possible. Stroop's (1935) results demonstrated that if the written word was a different colour than its printed colour (incongruent; e.g., the word ‘red’ written in green ink), participants reacted more slowly and less accurately. This is referred to as the Stroop Effect (MacLeod, 1991). Similarly, in trials in which the colour of the written word and its printed colour matched (congruent; e.g., the word ‘red’ written in red ink), participants reaction time and accuracy was improved. This is referred to as Stroop facilitation (MacLeod, 1991).

The Stoop task has been used to distinguish between top-down and bottom-up processing: colour naming was argued to be a controlled (or top-down) process, while word reading was argued to be an automated (bottom-up) process (MacLeod, 1992). According to conflict monitoring accounts (e.g. Botvinick et al., 2001) there is a conflict between the colour of the written word and its font for incongruent trials. The conflict triggers a control process which leads to the Stroop Effect. However, the exact nature of the conflict remains unclear (Algom & Chajut, 2019).

There are many variations of the Stroop task (for a comprehensive overview, please refer to MacLeod, 1991). One variation involves semantic manipulations. By manipulating the relation between word meaning and font colour, Klein (1964) demonstrated that words associated with a colour showed inferences, if presented in an incongruent condition. Specifically, the colour-related word (e.g. ‘lemon’) yielded higher response times if presented in a colour that did not match their relatedness (e.g. blue) compared to words that were unrelated to a colour (e.g. ‘friend’). Klein’s semantic Stroop Effect suggest that interference is largely automated.

Kinoshita and colleagues (2017) demonstrated in a Stroop-like task that colour naming for stings (e.g. ‘XXX’) was faster compared to words (e.g. ‘HAT’) and pseudo-words (e.g. ‘HIX’). This suggests that not only the meaning of a word but reading per se impacted reaction times in Stroop-like tasks. The authors argue that semantic processing in the Stroop task is goal-directed, insofar that only semantic features relevant to colour are evaluated. Hence, in Stroop-like tasks one differentiates between task conflict and information conflict:

task conflict refers to the conflict between the relevant task (i.e. identifying the font of a word) and the irrelevant task (i.e. reading the word). This type of conflict is evident in both congruent and incongruent trials. Second, *information conflict* refers to the conflicting information between font and meaning of a word, which is evident in incongruent trials (Littman, Keha & Kalanthroff, 2019).

Martino and Marks (1999) demonstrated Stroop-like interference in a cross-modal task using pitch and brightness. Higher pitch is perceived as brighter than lower pitch. The authors asked participants to identify high-pitch and low-pitch tones while viewing words on a screen. The words included brightness words (e.g. ‘black’ and ‘white’), brightness-related words (e.g. ‘night’ and ‘day’), words metaphorically related to brightness (e.g. ‘bad’ and ‘good’), non-linguistic words (including pseudowords, anagrams). These were presented in either black or white font. In line with the studies presented above, both brightness words and brightness-related words facilitated pitch processing in the congruent condition and interfered with processing in the incongruent condition. The authors found that pitch was identified fastest if pitch, word and font matched. This was the case for both brightness and brightness-related words (e.g. high pitch, ‘white’, white font and high pitch, ‘day’, white font). These findings support a cross-modal congruency effect for words related to brightness and pitch.

Stroop-like effects have also been used in studies of embodied cognition to investigate the connection between language and sensorimotor processing (e.g. Barsalou, 1999; Glenberg, 1997). For example, in different Stroop-like experiments Miller and Kaup (2020) presented participants with hand or foot related action words (e.g. grab, or kick), or non-words. This was followed by a response cue, indicating whether a hand or foot response (pressing a key on the keyboard, or pressing a pedal under the right foot) should be made. The results yielded a small congruency effect, such that for example a hand response was faster, if preceded by a hand-related action word. The effect was stronger, when participants were asked to withhold their responses if the stimulus was a non-word. This modification ensured that participants processed the words to the extent of identifying them as words or non-words, and thus magnifying information conflict. The results are in line with models of embodied cognition.

Taken together, congruency studies such as the Stroop Effect may shed light on language processing. Longer reaction times in incongruent conditions for semantically colour-related words suggest automaticity in language processing. Other studies have suggested attentional processes to be involves as well. There is evidence that this effect is stable across modalities, including semantic brightness and pitch as well as action words and action. These studies are in line with embodied theories of language.

Negations and Cognitive Control

Negations are semantic operations, which reverse the meaning of its positive counterpart. For example, the sentences 'Mary loves him' and 'Mary does *not* love him' express the same state of affairs, except that in the latter case it does not hold (Deprez & Espinal, 2020). In psycholinguistic research a key question is how negations are processed. Behavioural paradigms have robustly shown that negative sentences are processed more slowly and are more error-prone in comparison to positive sentences. Evidence suggests that negation processing involves a suppression mechanism, which reduces the availability of some presented information. This is referred to as the *negation effect* (for a comprehensive overview please refer to Kaup & Dudschig, 2020a).

It is less clear at what stage in processing this suppression mechanism is activated. According to two-step accounts, the processing of a negated sentence such as 'Tom does not wear a hat' involves two steps: first, the representation of the non-factual state of affairs (Tom with a hat); second, the representation of the factual state of affairs (Tom without a hat). On the other hand, one-step accounts such as the fusion model (e.g. Mayo, Schul & Burnstein, 2004) suggest that the factual state of affairs is represented right away (Kaup & Dudschig, 2020a). Evidence for the latter comes from neuroimaging studies, which have demonstrated a significant decrease in BOLD response for negative (e.g. ‘John has not left’) compared to affirmative sentences (e.g. ‘John has left’). In this case the affirmative sentence evokes the representation of John leaving through recruitment of the cortical action-representation system. In its negated form the same system may be at least partially suppressed (Tettamanti et al., 2008). Similarly, Foroni and Semin (2013) showed that negated sentences such as ‘I am not smiling’ result in below baseline activations of the same regions triggered both by their affirmative counterparts and the action itself (i.e. smiling). Both studies support the view of embodied language.

Dudschig and Kaup (2018) combined negation research with conflict research. Participants performed a spatial decision task: they had to react appropriately to the phrases ‘now left’, ‘not left’, ‘now right’ and ‘not right’. Behavioural results demonstrated that participants reaction time was slower, and their accuracy was worse when the phrase was negated. EEG analysis indicates a two-step process for negations: initially, the phrase ‘not right’ triggered the motor cortex for ‘right’ (i.e. the left motor cortex), then the contralateral-motor cortex was activated, reflecting the switch from ‘not right’ to left respond. The results are in line with paradigms studying cognitive conflict, such as the Stroop task, indicating that in case of negations both the to-be-negated and the factual state of affairs is activated. In Dudschig and Kaup’s (2018) study instructional stimuli are used to initiate an action (e.g. for ‘not left’ participants had to press the right key). Thus, no conclusions about the processing of negations for phrases where stimuli and responses are not explicitly matched can be drawn.

Taken together, the presented fMRI studies are in line with the one-step account when processing negations suggesting the activation of a ‘suppression mechanism’. Contrary the results of EEG study by Dudschig and Kaup (2018), which indicate a reverse in neural signature. This in line with the two-step account when processing negations. Similar to the Stroop task in which incongruent trials increase reaction times compared to congruent trials, negations appear to be processed slower in comparison their affirmative counterparts. Neither conflict research, nor negation research can provide a comprehensive account as to the nature of conflict and when it takes place during information processing.

Purpose of the Present Study

In line with Dudschig and Kaup (2018) we aim to reconcile conflict research with negation research, and to answer some of the outstanding questions from both fields. We used a semantic congruency paradigm in which congruency was masked. This means that participants were instructed to distinguish words based on a dimension different to that of interest. Explicitly, we asked participants to press one key for animal-related words and another key for other words (consisting of brightness-related, darkness-related, and control words). All words were displayed randomly in either white or black font resulting in congruent (e.g., the word ‘sun’ in white font) and incongruent (e.g., the word ‘sun’ in black font) conditions for the brightness-, and darkness-related words. In line with semantic Stroop Effect literature (e.g. Klein, 1964; Martino & Marks, 1999; Miller & Kaup, 2020) we expect reaction times for incongruent trials to be longer and less accurate in comparison to congruent trials.

Our experiment consists of two blocks, one with negations and one without negations, for which the block-order was counterbalanced. In other words, the same stimuli are randomly shown to the participants with and without negations. In line with the two-step account in negation processing and Dudschig & Kaup, (2018), we expect an increase in overall response times in the negated block compared to the block without negations, irrespective of block order. That is, for ‘no sun’ first ‘sun’ and then the negation will be processed.

Both increased response times for incongruent and negated trials respectively can be described in terms of conflict. Yet, this conflict may be resolved for incongruent negated trials. We expect an interaction between congruency and negations. That is, negated brightness-related stimuli (e.g. 'no sun') in the incongruent condition (e.g. white font) are processed similar to the two-step account: first the affirmative counterpart (e.g. 'sun') and the font colour (e.g. 'white') are processed; second the negation is processed. We expect that participants respond after the first step of processing, leading to faster response times for incongruent negated trials. Conversely, reaction times are predicted to be higher for incongruent and lower congruent trials in the block without negations.

Methods

Participants

We recruited 14 first year psychology students from the University of Groningen. Participants were all native Dutch speakers, who were compensated with course credits (1.1 SONA credits) for part-taking in the study. Additionally, we recruited 50 participants using the paid participant platform Prolific (www.prolific.co). For taking part in our experiment, participants received 1.88£. We included two pre-screening criteria, namely Dutch native speakers and participants could not sign up more than once to our study. The experiment was the same for both pools. The experiment was approved by the ethics committee of the University of Groningen.

Stimuli and Design

The stimulus set was adapted from Mathôt and colleagues (2019). The set can be categorised into four word categories: words related to brightness (e.g. 'zon', or sun; n = 25), words related to darkness (e.g. 'nacht', or night; n = 24), control words (e.g. 'jas', or jacket; n = 26) and animal names (e.g. 'muis', or mouse; n = 19). For the purpose of this study we extended the stimuli set as follows: we formed the corresponding negations to all stimuli. The negations and the single words constituted the two blocks of the experiment. In Dutch, nouns are negated using 'geen', while adjectives and verbs are negated using 'niet'. In the original list, the categories bright, dark and control consist of a mix of nouns, verbs and adjectives, while the category animal consists of animal names only. Hence, in the negation block ‘geen’ would be predictive of the category animal names and therefore of correct response. To exclude the possibility that negations were predictive of correct response we increased the stimuli related to animals to 80, consisting of animal names (e.g. ‘vis’, or fish), adjectives related to animals (e.g. ‘tam’, or tame) and verbs related to animals (e.g. ‘overwinteren’, or hibernate). Consequently, negations were similarly distributed in all word categories: about 60% of all word categories were negated with ‘geen’ and 40% were negated with ‘niet’. In total 312 stimuli were used during the experiment, 156 with and 156 without negations.

In line with Mathôt et al. (2019), we set the background screen to grey (#777777). The stimuli appeared on the screen in both a random order, as well as randomly in either black or white font (Table 1). In the single-word block, brightness-related words appearing in white font, and darkness-related words appearing in black font constitute the congruent condition. Correspondingly, brightness-related words appearing in black font, and darkness-related words appearing in white font constitute the incongruent condition. In the negation block, negated brightness-related words appearing in white font, and negated darkness-related words appearing in black font constitute the incongruent condition. Correspondingly, negated brightness-related words appearing in black font, and negated darkness-related words appearing in white font constitute the incongruent condition.

Block	Condition	Word Category	Font
Single word	Congruent	bright (e.g. sun)	white
		dark (e.g. night)	black
	Incongruent	bright (e.g. sun)	black
		dark (e.g. night)	white
Negation	Congruent	bright (e.g. not night)	white
		dark (e.g. no sun)	black
	Incongruent	bright (e.g. not night)	black
		dark (e.g. sun)	white

Table 1. Summary of conditions in the experiment.

Procedure

The experiment was implemented with the OSWeb version of OpenSesame 3.3 (Mathôt, Schreij & Theeuwes, 2012). Data collection was managed by JATOS (Lange, Kühn, & Filevich, 2015). Participants from the SONA pool received a unique link to the experiment, which expired after a single usage. Participants from the paid participant pool received a general link, which facilitated participants to access the experiment multiple times, but only complete it once. To facilitate keeping track of who participated in our experiment, we asked participants from the SONA pool to indicate their SONA number and participants from prolific to provide the last three numbers of their Prolific ID prior to the start of the experiment. Afterwards participants gave their informed consent to take part in the experiment.

Participants were instructed to press a key when an animal-related word appeared on the screen and press another key for all other words. We counterbalanced the key response, by randomly assigning participants one of the following subject numbers: 1, 2, 3, 4. Participants (n = 33), who obtained an even subject number (i.e. 2 or 4) responded to animal related words with the left arrow and other words with the right arrow. Participants (n = 30), who obtained an odd subject number (i.e. 1 or 3) responded to animal related words with the right arrow, and other words with the left arrow. The experiment started with a short practice trial, including 10 stimuli that were either animal related or control words, half of them were negated while the other half was not.

The experiment was divided into two blocks, negated and single word block. The block order was counterbalanced between participants. Participants (n = 29), who were randomly assigned subject number 1 or 2, saw the single word block first. Participants (n = 34), who were randomly assigned subject number 3 or 4, saw the negated block first. Each block consisted of 156 trials. Between the two blocks, participants were able to take a self-paced break. Each trial consisted of a fixation dot, which was shown to the participant for

500ms. This was followed by the stimulus, that is, a single word from one of the four categories. Stimuli were displayed randomly in black or white font colour, and their order was randomised. Stimuli were presented for 1500ms, or until the participants' key response. No responses were considered incorrect. Lastly, participants obtained feedback on their responses: a green fixation dot if the participant's key response to the word category matched the key response of the instructions. For example, the participant was instructed to respond with the left arrow to animal-related words, and she pressed the left arrow when the word *tam* (or tame) appeared on the screen. A red fixation appeared if the participant's key response to the word category did not match the key response of the instructions and when participants failed to respond within 1500ms. The experiment took on average 12:42min with a standard deviation of 2:42min.

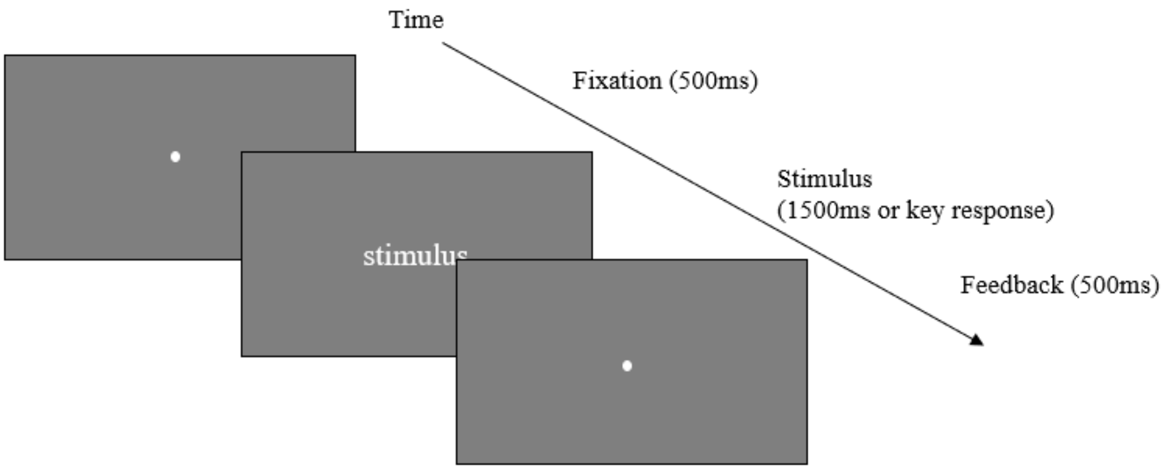

Figure 1. Trial sequence. Stimuli were randomly drawn from the four different word categories, randomly presented in black or white font. Compare Table 1 for conditions.

Results

Data Pre-processing

The pre-processing of the data included several steps: first, practice trials were excluded from the final analysis. Second, we excluded two participants, whose overall accuracy was two standard deviations below participants' mean accuracy. This resulted in a total of 62 participants. Third, we excluded all incorrect trials. As participants' mean response times were normally distributed, no further participants were excluded from the analysis. Fourth, we excluded trials with a response time below 300ms. We did not include a high cut-off score, as all trials that exceeded the set time-out of 1500ms were coded as 'incorrect' and therefore excluded from the analysis. Overall, the mean accuracy of the experiment was 87.23%. Response times were normally distributed with a mean of M = 689.694ms and a standard deviation SD = 182.892ms. Finally, for the congruency analysis and the interaction analysis we included congruent (i.e. brightness related words in white font and darkness related words in black font) and incongruent trials (e.g. darkness related words in white font and brightness related words in black font). This resulted in 5449 trials. For the negation hypothesis we also included animal and control words resulting in 17109 trials.

Main Analyses

For the main analyses we conducted a linear-mixed effects model (LME) using the lme4 package (Version X; Bates, Maechler, Bolker, & Walker, 2016) for the R software environment (Version X; R Development Core Team, 2016). The dependent variable was reaction time. We used congruency, negation as well as their interaction as fixed effects, and included by-participant random intercepts for the simplest models. Additionally, more complex models included random slopes for all fixed effects. We used a likelihood ratio test to compare the models.

Assumption Checks. Before starting the main analyses, we checked the assumptions underlying LMEs. First, the assumption of independent observations between participants was satisfied, as the data was derived from different participants independent of one another. Moreover, LMEs account for dependency of observations within participants (Spieler & Schumacher, 2019). Second, linearity was assessed by plotting the model residuals against the predictor and found to be satisfied. Third, we assessed whether there is equal variance of the residuals across subjects. Both a residual plot and the Levene's test suggests a violation of the assumption (F(1) = 7.702, p = .006). Lastly, a Q-Q plot suggests that the data is not normally distributed. We applied a logarithmic transformation to assess whether this would improve the relative normality of the data. This resulted in slight improvements with respect to homoscedasticity and normality but led to a gross violation against linearity. Ultimately, we decided to carry out the analyses without transformations.

No Difference in Reaction Times Between Congruent and Incongruent Trials. Visual inspection of the point-plot in Figure 2 suggests that there were no differences between congruent and incongruent trials. The LMEs supported this conclusion (Table 2): model 1 denotes the simple model with by-participants as random intercept and model 2 includes random slopes for congruency between participants. The intercept denotes participants average response times during incongruent trials (Model 1 *Estimate* = 710.575ms; Model 2 *Estimate* = 704.576ms). Model 1 predicts that congruent trials were processed 2.217ms slower that incongruent ones, whereas model 2 predicts a slowing of 3.004ms. Both models yielded no significant differences between congruent and incongruent trials (Model 1: t (5388.666) = 0.512, p = .608; Model 2: t (53.572) = 0.648, p = .594). The likelihood ratio test suggests that model 2 did not improve model fit ($\chi 2(2)$ = .779, p=.787).

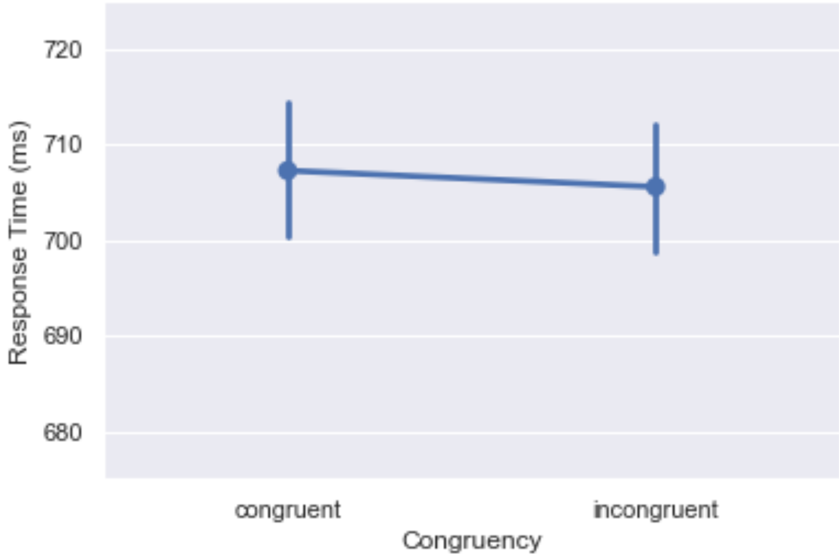

Figure 2. Point-plot Congruency. Differences in response times between congruent and incongruent trials

		Estimate	Std. Error	df	t value	p value
Model 1	Intercept	710.575	13.131	63.081	54.114	<0.0001***
	Congruent	2.217	4.326	5388.666	0.512	0.608
Model 2	Intercept	704.576	14.201	52.503	49.614	<0.0001***
	Congruent	3.004	4.636	53.572	0.648	0.594

Table 2. Linear Mixed Effect Models Congruency.

Reaction Times in Negated Blocks are Higher Compared to Blocks Without Negations. Visual inspection of the point-plot in Figure 3 suggests that there were differences between trials with and without negations. The LMEs supported this conclusion (Table 3): model 1 denotes the simple model with by-participants as random intercept and model 2 includes random slopes for negations between participants. The intercept denotes participants average response times during trials without negation (Model 1: Estimate = 686.194ms; Model 2 Estimate = 686.38ms). Model 1 predicts that negated trials were processed 15.074ms slower that incongruent ones, whereas model 2 predicts a slowing of 15.03ms. Both models yielded a significant difference between trials with and without negation (Model 1: t (17047.581) = 6.173, $p < .0001$; Model 2: t (59.86) = 2.489, $p = 0.016$). Importantly, block order did not affect the results (*F*(1) = .039, *p* =.844). The likelihood ratio test suggests that model 2 significantly improved model fit ($\chi 2(2) = 200.5$, $p < .001$).

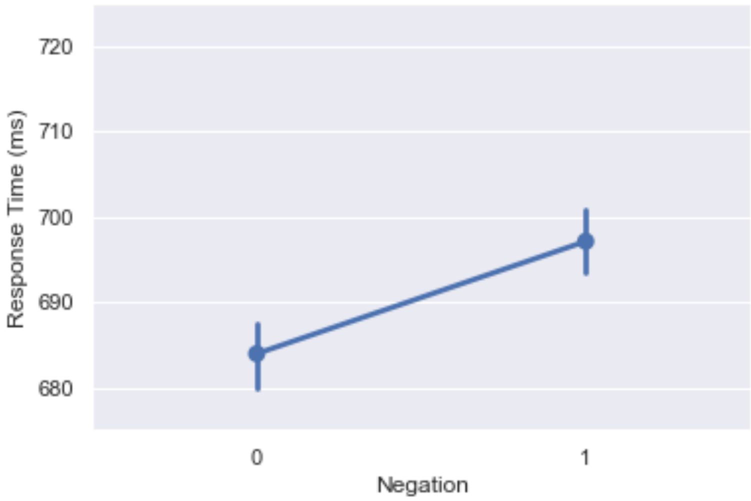

Figure 3. Point-plot Negations. Differences in response times between negated (here denoted as 1) and single word trials (here denoted as 0)

		Estimate	Std. Error	df	t value	p value
Model 1	Intercept	686.194	11.749	61.179	58.406	<0.0001***
	Negation	15.074	2.442	17047.581	6.173	<0.0001***
Model 2	Intercept	686.38	11.36	59.87	60.418	<0.0001***
	Negation	15.03	6.04	59.86	2.489	0.0156*

Table 3. Linear Mixed Effect Models Negations.

No Interaction Between Congruency and Negation. Visual inspection of the point-plot in Figure 4 suggests that congruent trials are processed slower if they are negated compare to trials without negations. No difference in response times between trials with and without negation can be detected for incongruent trials. The likelihood ratio test suggests that model 2, which included a random by-participant intercept and a random slope effect for negations between participants best fitted the data ($\chi 2(2) = 200.5$, $p < .001$). Table 3 presents the results from the LME analysis of model 2. The intercept denotes participants average response times during incongruent not negated trials (*Estimate* = 708.592ms). The model predicts that congruent trials without negations were processed 2.139ms faster than incongruent trials without negations. Incongruent negated trials were processed 4.834ms slower compared to their not negated counterpart. Finally, the interaction between congruency and negation suggests that congruent negated trials were processed 8.434ms slower than incongruent trials without negation. However, neither of these effects was significant (Congruency: t (5352.691) = - .356, p < .722; Negation: t (92.505) = .506, p = .614; Congruency*Negation: t (5344.828) = .986, p = .324).

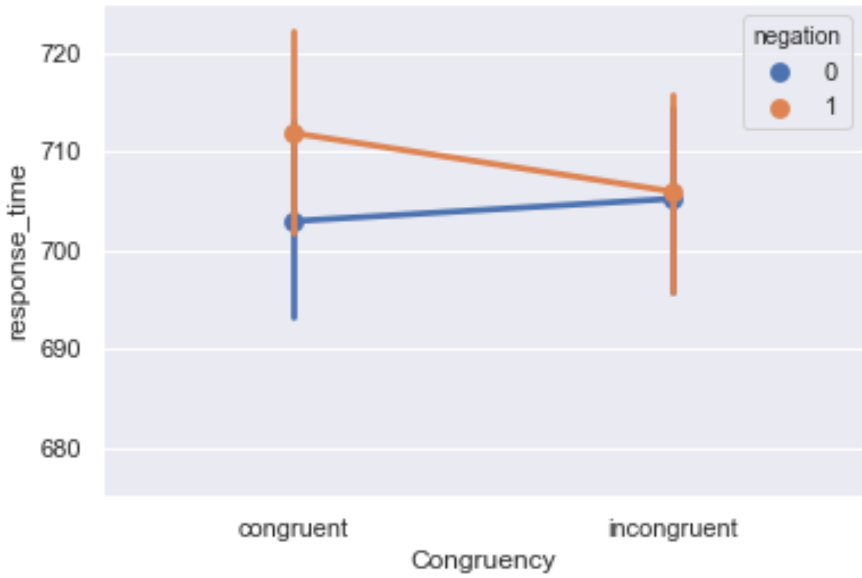

Figure 3. Point-plot. Interaction between Congruency and Negations.

		Estimate	Std. Error	df	t value	p value
Model 2	Intercept	708.592	12.996	66.772	54.524	<0.001***
	Congruent	-2.139	6.01	5352.691	-0.356	0.722
	Negation	4.834	9.546	92.505	0.506	0.614
	Congruent*Negation	8.434	8.553	5344.828	0.986	0.324

Table 3. Linear Mixed Effect Model. Interaction between Congruency and Negations.

Normative Ratings

For all brightness and darkness-related words, as well as their corresponding negations, we collected normative ratings from 28 first-year psychology students of the University of Groningen. Participants were all native Dutch speakers, who were recruited using the SONA platform. For their participation, we awarded participants with 1.4 SONA credits. The study was implemented using Qualtrics (Version X; Qualtrics, Provo, UT). After obtaining participants' informed consent, participants were asked to rate the brightness and darkness related words on a scale ranging from 0 to 100. On this scale, we asked participants to indicate the word's sense of brightness (from *very dark* to *very bright*) and the word's valence (from *very negative* to *very positive*). We presented six words at a time alongside with the rating scale. Brightness and valence were rated in separate, counterbalanced blocks. The data was analysed using JASP (version 0.12.2, JASP Team, 2018).

For the analyses we conducted four Bayesian ANOVAs and one Bayesian correlation. The first Bayesian ANOVA suggests that brightness related words are rated as brighter than darkness related words (BF >100, error < .001). The second Bayesian ANOVA suggests that brightness related words are viewed more positive than darkness related words (BF >100, error < .001). The third Bayesian ANOVA suggests that negated brightness related words are viewed darker than negated darkness related words (BF >100, error < .001). The fourth Bayesian ANOVA suggests that negated brightness related words are viewed less positive than negated darkness related words (BF >100, error < .001). We replicated the findings of Mathôt et al. (2019) and Mathôt et al. (2017) using a Bayesian correlation analysis. Specifically, we found a positive correlation between semantic brightness and valence (r = .803, BF >100), which suggests that words related to brightness are perceived more positive than words related to darkness. We found the similar pattern for negations. Specifically, we found a negative correlation between negated bright words and valence (r = -.174, BF >100), which suggests that negating a brightness related word is viewed more negative than negating a darkness-related word. The normative ratings analyses indicate that the words chosen were appropriate.

Discussion

The present study investigated the effect of syntax in language processing. Specifically, we investigated the effect of negations in a masked semantic congruency paradigm. Congruent trials were defined as trials in which the stimuli's meaning and font matched (i.e. the word 'sun' in white font, the phrase 'no sun' in black font). Incongruent trials were defined as trials in which the stimuli's meaning and font did not match (i.e. the word 'sun' in black font, the phrase 'no sun' in white font). Our results show that semantic congruency had no effect on participants' response time. While we found that participants' response times increased for negated trials, negations did not affect response times of congruent versus incongruent trials.

Our first hypothesis was that words related to brightness were processed faster if font colour corresponded to the meaning of the word. For example, the word 'sun' would be processed faster if presented in white font, rather than black font. Similarly, the phrase 'no sun' would be processed faster if presented in dark font rather than white font. We did not find differences in response times between the congruent and incongruent trials. Even when we consider congruent trials in the block without negations only (Figure 4, blue line), congruent trials were not processed faster than incongruent ones. In the following we consider stimulus set and task load as possible explanations.

Firstly, the absence of a congruency effect may be explained by our stimulus set. In his integrative review on the Stroop Effect, MacLeod (1991) concluded that the higher the semantic association between a word and its font, the higher its potential interference in incongruent conditions (MacLeod, 1991). Our stimulus set was based on the set of Mathôt and colleagues (2019), which demonstrated that brightness-related words used evoked pupil constriction and darkness-related words evoked pupil dilation. Together with the results of our normative rating analyses, we conclude that our stimulus set was high in semantic association.

Secondly, by setting the task to distinguish between animal-related words and other words, we masked our underlying research question and isolated information conflict, similarly to Martino and Marks (1999). While Martino and Marks (1999) used a cross-modal task, namely brightness (-related) words and pitch, we used a unimodal reaction time task.

Specifically, in Martino and Marks' (1999) experiment the congruent trials (word brightness matched font) facilitated pitch processing, whereas in our experiment congruency did not affect how quickly participants categorised words as animal-related or other. In line with cognitive load theory (Lavie & Dalton, 2014), we suggest that the cognitive load in our unimodal reaction time task exceeded the threshold for parallel processing. In our task, the differentiation between animal-related words and other words exhausted cognitive load, while the distractor in form of congruency was either rejected due to high cognitive load, and/or was too weak to cause interference. Overall, the results of our experiment suggest that semantic congruency does not affect reaction time.

Our second hypothesis was that in the negated block reaction times are slower compared to the block without negations. Our findings contribute to the existing body of evidence (e.g. Kaup & Dudschig, 2020b) that response times for trials with negations are slower compared to trials without negations. In the following we consider reading time, valence and negation effect as possible explanations:

Firstly, the slower reaction time in the negated block can be attributed to reading time. In our experiment the negated block consisted of two words, namely the negation and a second word. The block without negations consisted of a single word, only. It takes longer for a person to read two words compared to one word. For example, Graesser, and colleagues (1980) reported the average reading time per word to be 150-400ms. Our analysis revealed that the slowing in the negation block was on average 15ms. Hence, it is questionable, whether the slowing in the negated block can be attributed to reading time required for a second word.

Secondly, the density hypothesis (Unkelbach, Fiedler, Bayer, Stegmüller & Danner, 2008) postulates that positive information is processed faster. In line with this hypothesis, our normative ratings analyses demonstrate that brightness-related words are rated less positive when negated. Contrary, darkness-related words are rated more positive if negated.

Consequently, we consider valence as an explanation for slower reaction times in the negated block for unlikely.

Thirdly, the slower response time in the negated block can be attributed to the negation effect arising from cognitive conflict. The experiment of Dudschig and Kaup (2018) established conceptual and processing parallels between conflict and negation research. The authors suggest that the processing of negations can be regarded as a special type of information processing conflict. Accordingly, we argue that in our experiment, higher reaction times in negated trials may be ascribed to the cognitive conflict arising from the negation.

Our third hypothesis was that reaction times in the negated block would decrease for incongruent trials and increase for congruent trials. For example, the phrase ‘no sun’ would be responded to faster if presented in white font, compared to black font. While Figure 4 suggests such a trend, our analysis did not confirm an interaction between negations and congruency. As suggested earlier, the manipulation of congruency may not have been strong enough to interfere with the task at hand. We conclude that in our experiment reaction time in the negated block did not depend on congruency. Thus, we did not establish behavioural evidence for an embodiment of syntax.

The present study is unique in that it combines negation and conflict research, while trying to establish that if both types of conflict are present, they cancel each other out (hypothesis three). While the experiment of Dudschig and Kaup (2018), demonstrated that negation processing can be regarded as a special type of information processing conflict, the results of our study show that they should be differentiated. Even though the experiment was carried out online, the use of a mixed effects model in our analyses allowed us to account for violations against the assumption of independence (Singmann & Kellen, 2020). Together with a post hoc power analysis, which revealed the sufficiency of our sample size, the use of LMEs supports the reliability of our results.

We instructed participants to respond to animal related words with one key (left or right) irrespective of the block (with or without negation). This means, participants had to use the same key response for the ‘cat’ and ‘no cat’. Hence, it is possible that the negations were not processed and participants only paid attention to the second word. While the increased reaction time in the negated block suggests that these trials were processed differently, we cannot establish that this reflects the processing of the stimuli’s meaning. If anything, the absence of a congruency effect suggest that meaning may not have been processed beyond the task at hand. This issue should be addressed in future studies.

Our recommendation for future studies is threefold: control for the number of words, increasing saliency of congruency and counterbalancing key response within subjects. To control for the number of words in the different blocks, future research should balance word count between blocks. For example, Dudschig and Kaup (2018) used phrases such as ‘not left’ and ‘now left’ to account for reading time. In future experiments one could use the pronouns ‘much/many’ and the adverb ‘very’ in the block without negation. This would not only ensure equal word count between blocks, but moreover make congruency more salient. ‘Much sun’, or ‘very bright’ may be perceived as brighter than the words ‘sun’ and ‘bright’ alone. Moreover, counterbalancing key response within the experiment may promote the processing of meaning in the negated block. For instance, some participants would have to respond ‘right’ to animal related words and ‘left’ to the negations thereof, while others would have to respond ‘right’ on both occasions. This may promote the processing of negations.

In conclusion, the present study contributed to the existing literature by providing a novel approach to combine conflict and negation research. While we successfully replicated the effect of negations on reaction time, the results of our masked congruency paradigm suggest that semantic congruency does not affect participants’ reaction times. We have proposed that the conflict arising from negations should be distinguished from that arising of incongruency.

References

Algom, D., & Chajut, E. (2019). Reclaiming the Stroop Effect Back From Control to Input-Driven Attention and Perception. *Frontiers in Psychology*, *10*, 1683. https://doi.org/10.3389/fpsyg.2019.01683

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Ben Shalom, D., & Poeppel, D. (2008). Functional Anatomic Models of Language: Assembling the Pieces. *The Neuroscientist*, *14*(1), 119–127. https://doi.org/10.1177/1073858407305726

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Cancho, R. F. i, & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1482), 2261–2265. https://doi.org/10.1098/rspb.2001.1800

Dudschig, C., & Kaup, B. (2018). How does "not left" become "right"? Electrophysiological evidence for a dynamic conflict-bound negation processing account. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 716–728. https://doi.org/10.1037/xhp0000481

Foroni, F., & Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00209

Galetzka, C. (2017). The Story So Far: How Embodied Cognition Advances Our Understanding of Meaning-Making. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01315

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In *Creative thought: An investigation of conceptual structures and processes* (S. 495–522). American Psychological Association. https://doi.org/10.1037/10227-018

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, *19*(2), 135–151. https://doi.org/10.1016/S0022-5371(80)90132-2

Kaup, B., & Dudschig, C. (2020a). Understanding Negation: Issues in the processing of negation. In V. Déprez & M. T. Espinal (Hrsg.), *The Oxford Handbook of Negation* (S. 634–655). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198830528.013.33

Kaup, B., & Dudschig, C. (2020b). Understanding Negation: Issues in the processing of negation. In V. Déprez & M. T. Espinal (Hrsg.), *The Oxford Handbook of Negation* (S. 634–655). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198830528.013.33

Kinoshita, S., De Wit, B., & Norris, D. (2017). The Magic of Words Reconsidered: Investigating the Automaticity of Reading Color-Neutral Words in the Stroop Task. *Journal of experimental psychology. Learning, memory, and cognition*, *43*(3), 369–384. https://doi.org/10.1037/xlm0000311

Klein, G. S. (1964). Semantic Power Measured through the Interference of Words with Color-Naming. *The American Journal of Psychology*, *77*(4), 576–588. JSTOR. https://doi.org/10.2307/1420768

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J.B., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., et al. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. ELPUB.

Krause, F., & Lindemann, O. (2014). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, 46(2), 416–428.

http://doi.org/10.3758/s13428-013-0390-6

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), 1–14.

Lavie, N., & Dalton, P. (2014, Januar 1). *Load Theory of Attention and Cognitive Control*. The Oxford Handbook of Attention. https://doi.org/10.1093/oxfordhb/9780199675111.013.003

Littman, R., Keha, E., & Kalanthroff, E. (2019). Task Conflict and Task Control: A Mini-Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01598

MacLeod, C. M. (1991). Half a century of research on the Stroop Effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Martino, G., & Marks, L. E. (1999). Perceptual and Linguistic Interactions in Speeded Classification: Tests of the Semantic Coding Hypothesis. *Perception*, *28*(7), 903–923. https://doi.org/10.1068/p2866

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary Responses to Words That Convey a Sense of Brightness or Darkness. *Psychological Science*, *28*(8), 1116–1124. https://doi.org/10.1177/0956797617702699

Mathôt, S., Sundermann, L., & Rijn, H. van. (2019). The effect of semantic brightness on pupil size: A replication with Dutch words. *BioRxiv*, 689265. https://doi.org/10.1101/689265

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*(4), 433–449. https://doi.org/10.1016/j.jesp.2003.07.008

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Miller, J., & Kaup, B. (2020). Influences of task and attention on action verb congruence effects: How automatic are embodiment effects? *Acta Psychologica*, *210*, 103155. https://doi.org/10.1016/j.actpsy.2020.103155

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: http://www.R-project.org/.

Spieler, D., & Schumacher, E. (2019). *New Methods in Cognitive Psychology*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/rug/detail.action?docID=5965779

Stroop, J. R. (1935). *Studies of Interference in Serial Verbal Reactions*. 9.

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2008). Negation in the brain: Modulating action representations. *NeuroImage*, *43*(2), 358–367. https://doi.org/10.1016/j.neuroimage.2008.08.004

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*(1), 36–49. https://doi.org/10.1037/0022-3514.95.1.36

Willems, R. M., & Casasanto, D. (2011). Flexibility in Embodied Language Understanding. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00116

Appendix

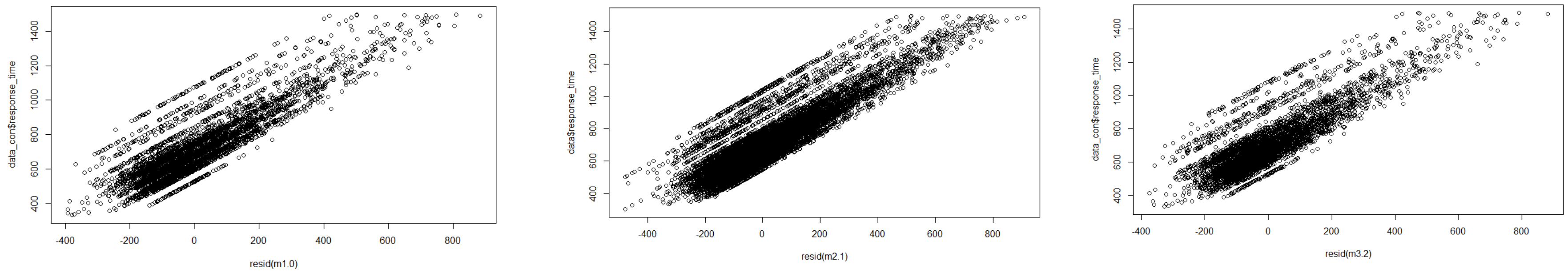

Figure A1. Linearity assumption for the three analyses: Congruency, Negation and the Interaction (left to right).

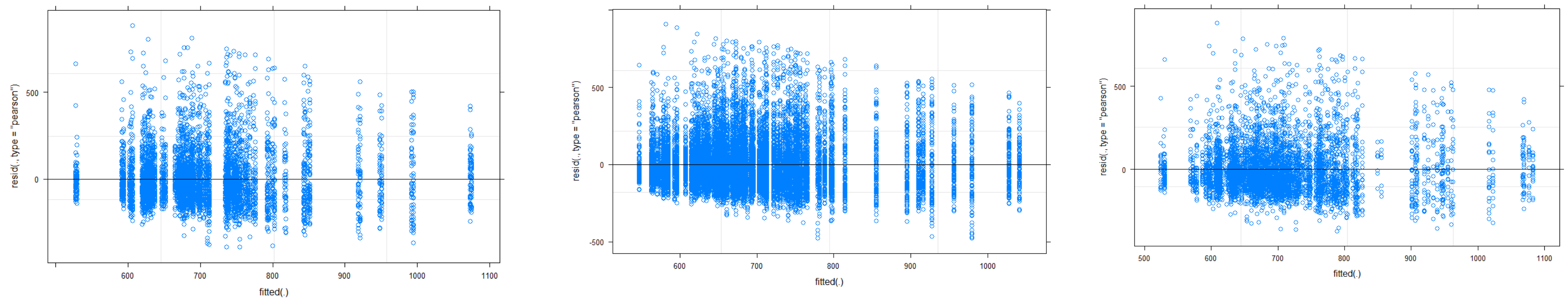

Figure A2. Homoscedasticity assumption for the three analyses: Congruency, Negation and the Interaction (left to right).

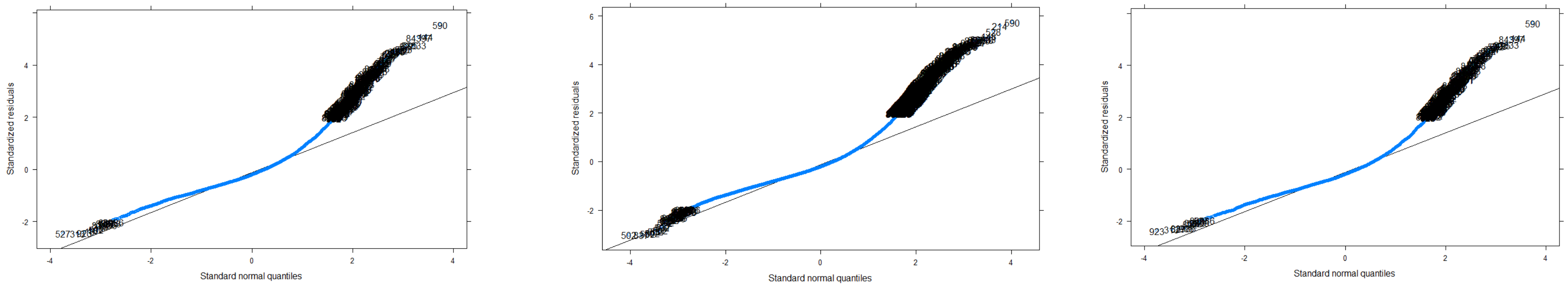

Figure A3. Normality assumption for the three analyses: Congruency, Negation and the Interaction (left to right).