Eberhard Karls Universität Tübingen

Seminar für Sprachwissenschaft

The Role of Metaphor and Reliability of the Speaker in Reasoning: A Replication Study

Research Apprenticeship

Winter term 2022/2023

Miriam Schiele

Matriculation number: 6098290

Supervised by

Prof. Dr. Michael Franke

Date of submission: April 17th, 2023

Table of contents

Table of figures

Abstract

The aim of the follow-up study was to examine whether the effect of metaphors on reasoning is influenced by the (underlying) assumptions of the participants. That is, the use of a particular metaphor is not seen as a random decision but shows that the speaker has reflected on the issue and concluded that this metaphor best describes the problem, allowing participants to propose appropriate solutions.

1. Introduction

Metaphors can be found in everyday speech. Their persuasive nature has been object of studies since the last 1970s (cf. Gibbs, 2010). For instance, Thibodeau & Boroditsky (2011) researched how even subtle metaphors impact decision-making regarding a socio-political issue, namely crime. In their study, they explored how metaphorically framing crime as a *beast* or as a *virus* influences the suggestions offered by participants to solve the issue. That is, when crime was metaphorically described as a *beast*, participants were more likely to suggest enforcement-oriented solutions to the issue than when crime was metaphorically framed as a *virus*.

So far, however, little research has been conducted regarding the question whether the reliability of the speaker influences the effect of the metaphor. Hence, the present study is divided into a replication study and a follow-up study. The replication study aims to replicate the experiment conducted by Thibodeau & Boroditsky (2011). The main hypothesis (1), therefore, holds that metaphorically framing crime as a *beast* triggers participants to prefer enforcement-oriented solutions and describing crime as a *virus* influences participants to favor reform-oriented solutions. Additionally, we test hypothesis (2) that participants' political affiliation influences their responses, and hypothesis (3) that metaphor-framing effect does not vary with vignette length.

The follow-up study, on the other hand, examines whether the effect of metaphors on reasoning is influenced by the (underlying) assumptions of the participants. That is, the use of a particular metaphor is not seen as a random decision but shows that the speaker has reflected on the issue and concluded that this metaphor best describes the problem, allowing participants to propose appropriate solutions. Unlike in the replication study, participants are not only presented with a metaphorical description of the crime issue but also with pictures of different speakers that are associated with different levels of reliability, namely a newscaster and an aggressive looking, drunk person. We hypothesize that compared to the *hooligan* condition, we expect a higher proportion of expected suggestions (reform-oriented suggestions in *virus* frame and enforce-oriented suggestions in *beast* frame) in the *newscaster* condition.

The paper is subdivided as follows. Chapter 2 outlines the research in the field of metaphors in reasoning. Chapter 3 focuses on the study design, including the pilot study, the replication study, and the follow-up study. In chapter 4, I will present the results and in chapter 5, I will discuss the findings. Finally, chapter 6 provides a summary of the study as well as an answer to the research question.

2. Literary review

In the following section, I will outline the research in the field of metaphors in reasoning. First, the scientific background of metaphors is explained, including the conditions under which metaphors are most influential. Second, I will summarize the study by Thibodeau & Boroditsky (2011) on metaphors in reasoning and discuss their results. Next, follow-up studies that could replicate the results by Thibodeau & Boroditsky (2011) as well as their further findings are recapitulated. Finally, I will present studies that could not replicate the findings and discuss their results.

Generally, metaphors can be described as "instances of non-literal language that involve some kind of [implicit] comparison or identification" (Knowles & Moon, 2005, p. 7). According to the conceptual metaphor theory, people speak and think in metaphors (cf. Thibodeau & Flusberg, in press). Consequently, recent studies have focused under which conditions and the degree to which metaphors can influence one's thoughts and decisions about socio-political topics such as crime (cf. Thibodeau, Hendricks & Boroditsky, 2017, Thibodeau, Matlock & Flusberg, 2019, Steen, Reijnierse & Burgers 2014, Thibodeau & Flusberg, in press, Reijnierse et al., 2015).

For metaphors to have their greatest impact, several factors must be aligned. First, the source domain (e. g. *virus* or *beast*) and target domains (e. g. *crime*) must match for the metaphor to be apt. Second, people must not already have strong beliefs about the topic in question. Third, prior knowledge of the topic as well as knowledge and interest in the source domain is useful for the metaphor to be persuading. Next, the entailment of metaphors also depends on factors other than the source domain. That is, the meaning of common metaphors might vary. Finally, metaphors are more influential, if the metaphorical mapping is extended in ways that align with

the metaphor. For instance, if the language used to describe the metaphor align with the mapping, people more likely to favor the metaphor-consistent response (cf. Thibodeau, Hendricks & Boroditsky, 2016).

However, little research has been conducted so far regarding the question on the influence of the speaker of the metaphor. That is, whether the reliability of the speaker corelates with the extent to which the metaphor is influential. For instance, if the speaker seems less reliable, the hearer is less susceptible to the metaphor.

In the original study that is to be replicated in the present paper, Thibodeau & Boroditsky (2011) found that metaphors can heavily influence reasoning. In their first experiment, 485 participants were given a short metaphorically framed report about increasing crime rates in the fictitious City of Addison, which included crime statistics of said city. For half of the participants, crime was described as *virus* infecting the city, while for the other half, as a *beast* attacking the city. Afterwards, participants were asked to propose solutions for the crime problem. Participants that were confronted with the metaphor of crime as a *virus* were more likely to propose dealing with the underlying problem and suggesting reforms to prevent crimes. Contrary, metaphorically framing crime as a *beast* led participants to focus on police force or other methods of law enforcement. Interestingly, participants identified the crime statistics in the report as the motivation for their problem-solving decision instead of the metaphor, although the crime statistics were identical in both vignettes. Thus, metaphors seem to subconsciously influence participants' decisions. Additionally, their results suggest that overall, Republicans are more likely to propose enforcement solutions.

In a replication study, Thibodeau & Boroditsky (2013) asked participants to evaluate a set of problem-solving solutions for the crime problem and select the most effective ones. The results were consistent with those of the original study. Similarly, Thibodeau & Boroditsky (2015), Thibodeau, Iyiewaure & Boroditsky (2015), and Thibodeau (2016) could replicate the findings. However, Thibodeau & Boroditsky (2015) found indications of a cultural shift in the opinion on the topic of crime. That is, participants' suggestions in their study in 2015 were focused on community outreach, while

in the original study in 2011, they proposed "policies grounded in the economy and prison system" (p. 11). Additionally, neighborhood watches were no longer seen as reform-oriented and consistent with the *beast* frame, as it was the case in 2011, but were seen as enforce-oriented instead (cf. Thibodeau & Boroditsky, 2015). Moreover, their results suggest that metaphors are more influential if the metaphoric framing is extended into the texts of policies. That is, participants were more likely to suggest social reforms that were labelled as *treatments* in response to a crime *virus*, while they favored *attacking* the problem with tough enforcement in response to a crime *beast* (cf. Thibodeau, 2016). As in the original study, results of both replication studies showed Republicans, overall, were more likely to be enforcement-oriented (cf. Thibodeau & Boroditsky, 2015, Thibodeau, 2016).

Likewise, Christmann & Göhring (2016) could replicate the findings as well. However, participants were Germans instead of US citizens, as it was the case in the original study by Thibodeau & Boroditsky (2011). This might have led to different results, since "the use and evaluation of specific metaphors can vary across cultures" (Steen, Reijnierse & Burgers, 2014, p. 7). Additionally, the coding of the answers differed, insofar as that answers that included an equal number of suggestions for enforcement and reform were not counted to either enforcement or reform, respectively, by 0.5, as it was the case in Thibodeau & Boroditsky (2011) but were coded as 0 instead (cf. Christmann & Göhring 2016). Interestingly, when conducting the coding in the same way as in the original study, the difference between the responses of the *beast* and the *virus* group failed to reach statistical significance (cf. ibid.).

The following two studies by Steen, Reijnierse & Burgers (2014) and Reijnierse et al. (2015) could not replicate the findings of the original study either. The results by Steen, Reijnierse & Burgers (2014) failed to show statistically significant differences between the metaphorically frames but suggested that enforcement solutions are preferred overall. However, as pointed out by Thibodeau & Boroditsky (2015), the coding of the answer differed from the original study in a way that is psychologically and statistically problematic. Moreover, only their fourth experiment included a sufficient sample. According to Thibodeau & Boroditsky (ibid.), it would be more

appropriate to treat the data as categorial. In their categorical data analysis, they found that participants of the *virus* group favored reform-oriented options that are consistent with the *virus* frame (cf. Thibodeau & Boroditsky, 2015).

Furthermore, in their series of replication studies, Reijnierse et al. (2015) could not find similar evidence as in the original study by Thibodeau & Boroditsky (2011). That is, they found no significant effect for the metaphorical framing on the preferred measures (cf. Reijnierse et al., 2015). However, instead of suggestion solutions, participants were asked to evaluate the effectiveness of two sets of measures on a scale and the presented text of the crime problem severely differed from the one in the original study. Not only did they include the mayor of the city as speaker but also information about the longevity of the crime issue, the length of the report, and the content of the report itself differed. That is, the report did not include statistical information and read *Crime is a virus/beast* instead of *Crime is a virus infecting the city/wild beast preying on the city*, as it was the case in the original study. The authors admitted that the design of the study was inappropriate in such a way that it "may have made it impossible to find out whether people actually reason by working out the entailments of the metaphorical frame" (Reijnierse et al., 2015, p. 260). Yet, as in the original study, the effect that Republicans, overall, were significantly more likely to find the enforcement-oriented approaches more effective than Democrats and Independents could be replicated (cf. Reijnierse et al., 2015).

3. Replication study

The following section provides an overview of the replication study. The study design, including the pilot study, procedure, materials, and methods of data analysis, is explained. Additionally, socio-demographic information about participants and the results of the study are given.

3.1 Study design

The experiment was conducted using a 2 (metaphorical frame: *beast* or *virus*) x 2 (vignette length: short or long) between-subjects design. The four

options of the metaphorically framed description of crime are given in the appendix.

After a welcoming page, participants were randomly presented with one of the four possible, metaphorically framed crime descriptions about crime in the City of Addison. While the description was still visible on the screen, participants were asked to propose solutions for the crime problem in a text box. Next, they were asked to rate the reliability of the text on a scale from 1 to 7. Finally, they were asked to indicate their political affiliation, age, education, gender, nationality, native language, and level of education. During the experiment, participants were not able to return to the pages they had already visited in the survey and update their responses. The code for the study can be found here: https://tinyurl.com/mr4vky58.

Participants' solutions were treated as a set of suggestions. Each suggestion is classified as either *enforce* or *reform* by hand. As in the original study by Thibodeau & Boroditsky (2011), suggestions are categorized as *reform* if the proposed solution suggests investigating the underlying cause of the problem or suggests a particular social reform to treat or inoculate the community. In contrast, suggestions are categorized as *enforce* if the proposed solution focuses on the police force or other methods of law enforcement or modifying the criminal justice system. As pointed out by Thibodeau & Boroditsky (2015), suggestions of *neighborhood watches* cannot be clearly classified and are therefore excluded from the analysis. Solutions that contain the same number of suggestions for both reform and enforce are categorized as *both*. Finally, solutions are categorized as *neither* if the proposed solution lacked a suggestion and are therefore excluded from the analysis. Participants' answers and their categorizations can be found here: https://tinyurl.com/bdexsaz7. The remaining three categories are ordered as follows: *reform*, *both*, *enforce*.

As preregistered at https://osf.io/2tbj5/, data were analyzed using a Bayesian ordinal regression model for the binary choice variable `metaphor` as covariate (either *beast* or *virus*). The analysis uses the statistical programming language R and relies on the `brms` package. In the analysis, the default (flat) priors of the `brms` package for the effect coefficients are used. The code for the analysis can be found here: https://tinyurl.com/yczvyjy3.

The preregistered study design is based on a pilot study that was completed by a total of 60 participants. Participants' answers as well as the findings of the pilot study can be found here: https://tinyurl.com/9rj4ya2j.

As preregistered, we hypothesized that compared to the *virus* metaphor, we expect a higher propensity of suggestion in the category *enforce* in the *beast* frame, by comparing the posterior estimates for the aggregate value in the *beast* condition with those from the *virus* condition. We judge there to be positive evidence in favor of the main hypothesis (1) if the posterior probability of the difference between the ordinal predictor value for the *beast* metaphor and the *virus* metaphor being positive (*beast* bigger than *virus*) is at least 0.95.

Furthermore, the study aims to test the hypothesis (2) that rates of *enforce* increase for participants that identify their political affiliation as Republican. As preregistered, we judge there to be evidence in favor of the hypothesis, if the posterior probability of this difference being bigger than zero is at least 0.95. We intend to also investigate the hypothesis (3) that the effect of metaphors on reasoning, as proposed in hypothesis (1), does not vary with the length of the metaphorically framed description of crime. In accordance with the preregistration, we judge there to be evidence in favor of the hypothesis, if the posterior probability of this difference not being bigger than zero is at least 0.95.

3.2 Participants

The data used in this study was taken form Prolific (www.prolific.co/). To ensure trustworthy work, the approval rate on Prolific was set to 95%. During the set-up of the study, it was ensured that only participants who self-identified as native English speakers based in the US were recruited. Every participant was allowed to participate only once and was paid for their participation.

A total of 200 participants completed the study, of whom 92 (46.0%) identified as females, 102 (51.0%) as males, 3 (1.5%) as other and 3 (1.5%)

preferred not to say. Their ages range from 18 to 85[1] with a mean of 38.9. A total of 65 participants (32.5%) completed high school as their highest level of education, while 2 (1%) participants did not graduate from high school. 96 participants (48%) reported college as their highest education, and 33 participants (16.5%) had a higher degree, while 4 (2%) preferred not to say. A total of 108 (54%) participants described themselves as Democrats, 33 (16.5%) as Republicans, while 56 (28%) identified as neither and 3 participants (1.5%) preferred not to say. The full data set can be found here https://tinyurl.com/fu42mzce. However, because only 3 participants out of 200 participants preferred not to state their political affiliation, we excluded those participants from the analysis to conduct a reliable statistical analysis.

3.3 Findings

The aim of the replication study was to test whether metaphorically framing crime as a *beast* triggers participants to prefer enforcement-oriented solutions and describing crime as a *virus* influences participants to favor reform-oriented solutions, as has been claimed by Thibodeau & Boroditsky (2011). Additionally, the replication study explores the influence of the vignette length on the perceived reliability. Since the follow-up study tests the hypothesis that the effects of both the *beast* and the *virus* frame are higher if the speaker describing the crime issue seems reliable, it is crucial to ensure that it is not the vignette length but the perceived reliability of the speaker that amplifies such effects.

First, consider Figure 1, in which participants' responses to the crime issue are depicted. It becomes apparent that enforcement-oriented solutions are preferred overall.

---

[1] One participant reported an age of 1850. It can only be assumed that this data point was caused by a typing error. Therefore, the second highest value is given here as the highest age and the mean value of age is calculated without the outlier.
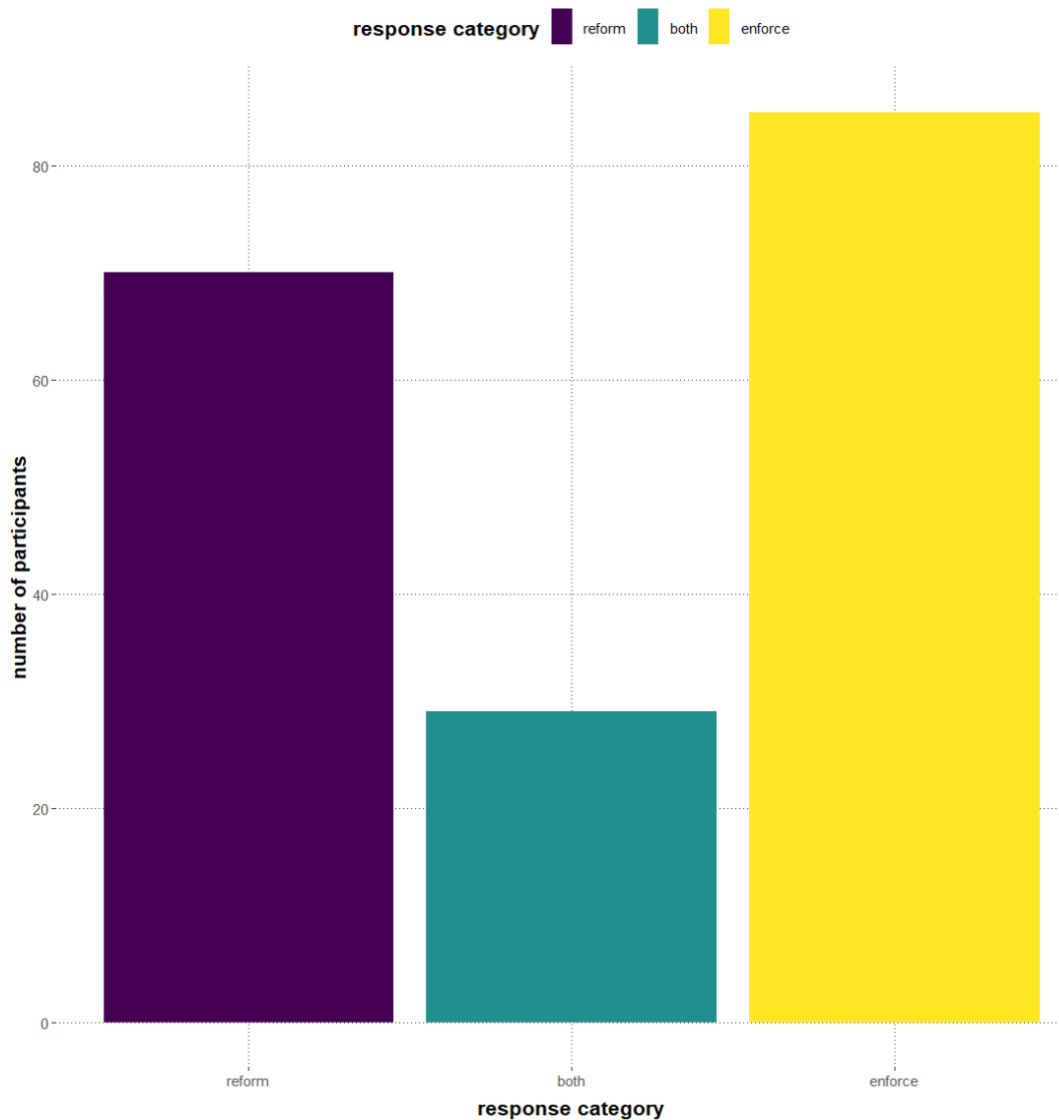
Figure 1: Categorizing participants' responses reveals that enforcement-oriented solutions are preferred overall.

Next, consider Figure 2, which presents the same data but differentiates between political affiliations and the metaphorical framing of the description. Considered jointly, it becomes apparent that on the one hand, Republican mostly suggest enforcement-oriented solutions overall and on the other hand, metaphorically framing crime as a *virus* leads participants to favor reform-oriented solutions.
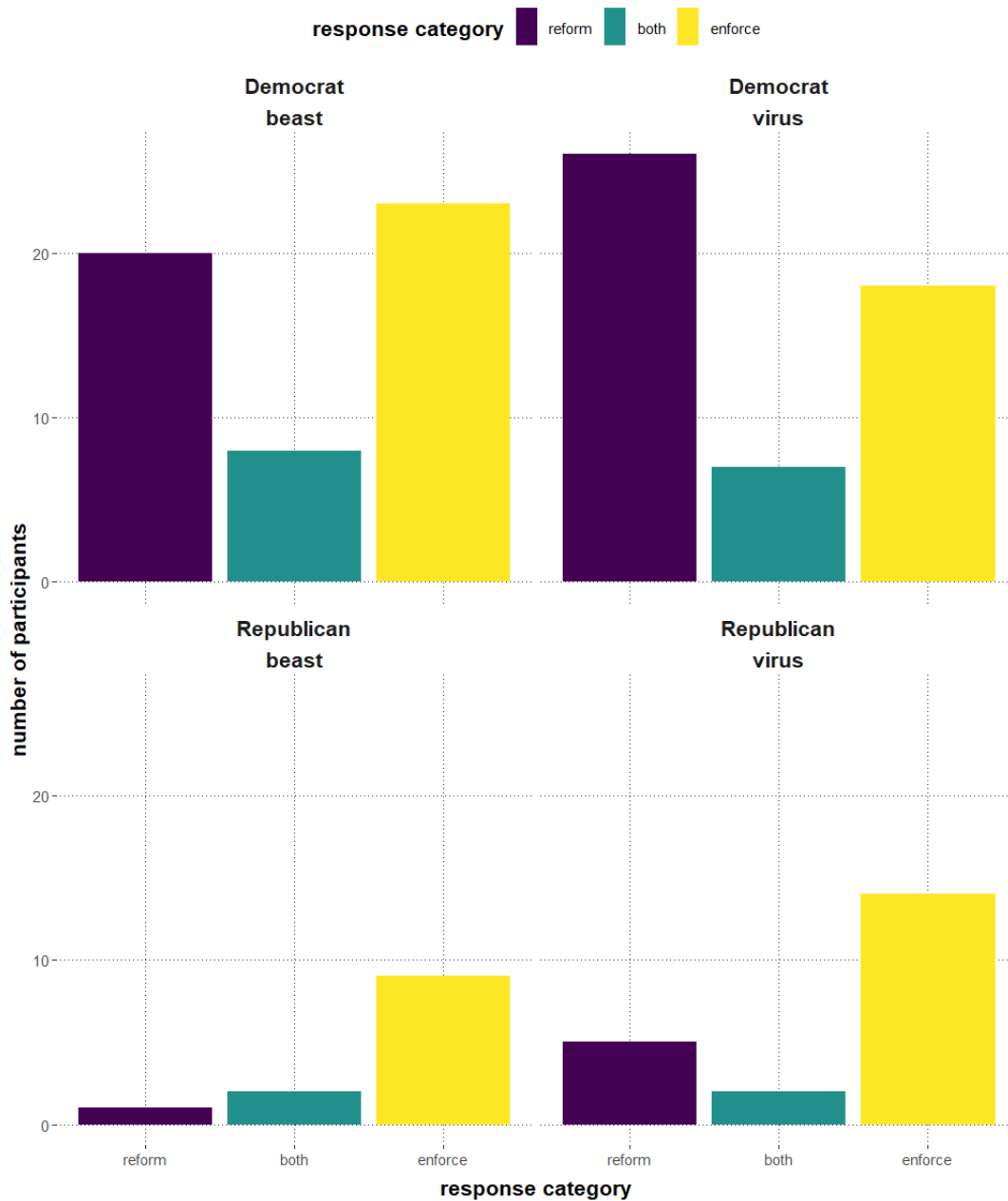
Figure 2: Participants' responses are influenced by the metaphorical framing of the description as well as their political affiliations.

As preregistered, we hypothesized that compared to the *virus* metaphor, we expect a higher propensity of suggestion in the category *enforce* in the *beast* frame, using Bayesian ordinal regression. We judge there to be positive evidence in favor of the main hypothesis if the posterior probability of the difference between the ordinal predictor value for the *beast* metaphor and the *virus* metaphor being positive (*beast* bigger than *virus*) is at least 0.95. However, the results of the replication study fail to reach the required

threshold, as the posterior probability is 0.67. Hence, the probability of hypothesis (1), given our model and data, is less than 0.95 and therefore insufficient to support the hypothesis.

Regarding the hypothesis (2) that investigates whether the rates of *enforce* increase for participants that identify their political affiliation as Republican, the results of the original study could be replicated. That is, as preregistered, we judge there to be evidence in favor of the hypothesis, if the posterior probability of this difference being bigger than zero is at least 0.95. The results revealed that the posterior probability is 0.99. Therefore, the results replicate the findings of the original study in this aspect.

Last, consider Figure 3, in which the influence of the vignette length on the perceived reliability is visualized. The mean values of the perceived reliabilities r are almost similar in both groups ($r = 4.8$ in the long vignette condition and $r = 5.1$ in the short vignette condition), although the results reveal that the variation is rather high ($\sigma = 1.5$ and $\sigma = 1.3$, respectively).
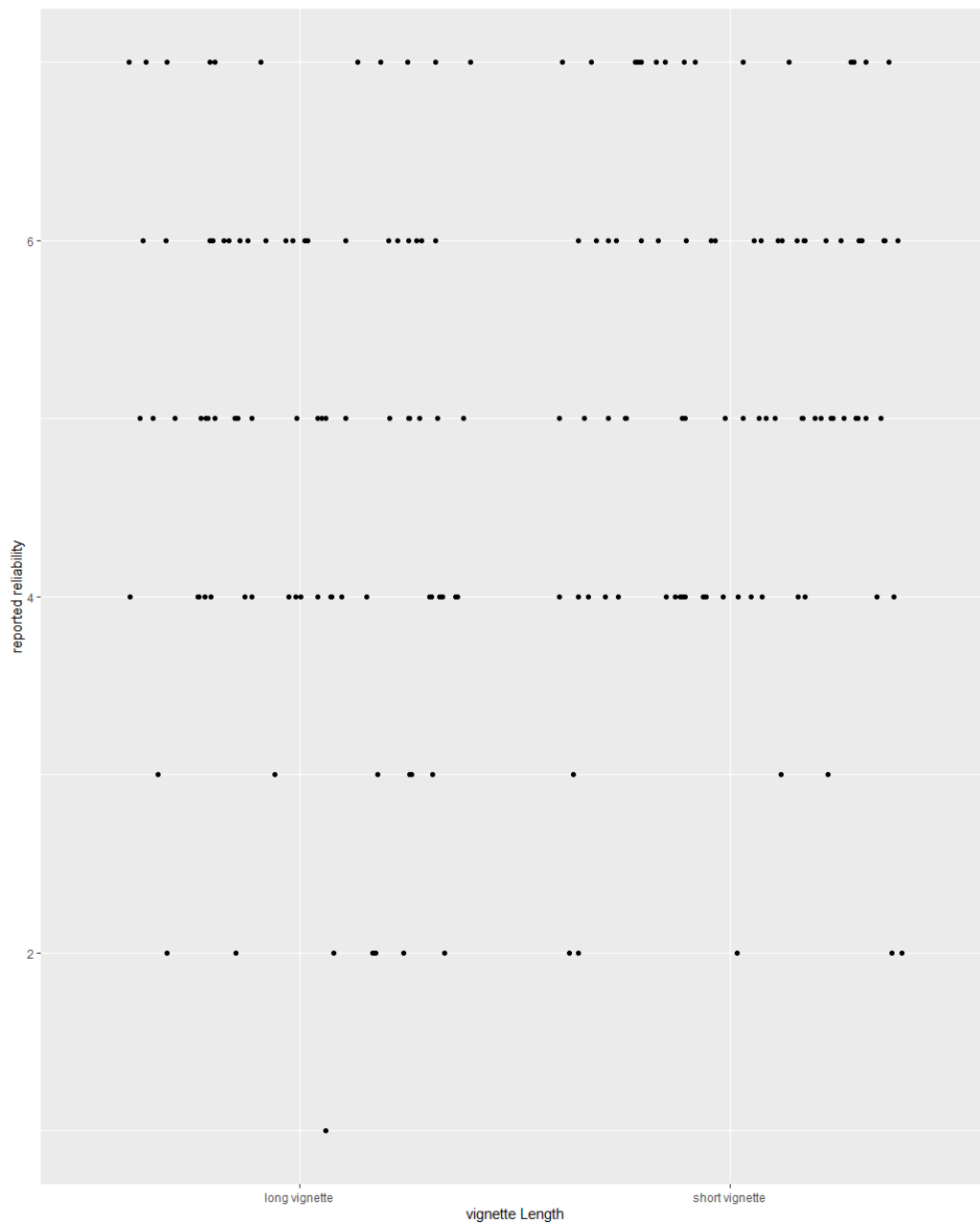
Figure 3: Plotting reliability against vignette length reveals hardly any influence of vignette length on perceived reliability.

As preregistered, we hypothesized that the effect of metaphors on reasoning, as proposed in hypothesis (1), does not vary with the length of the metaphorically framed description of crime. We judge there to be evidence against the hypothesis, that is, evidence in favor of a correlation between vignette length and the effect of metaphors on reasoning, if the posterior probability of the difference in vignette length being bigger than zero is at least 0.95. The results reveal that the posterior probability is about 0.82. Thus, there does not seem to be a correlation between vignette length and

the effect of metaphors on reasoning. Therefore, it can be assumed that the results support hypothesis (3).

4. Follow-up study

In the following, the study design, the participants, and the results of the follow-up study are presented.

4.1 Study design

Unlike the replication study, participants are not only presented with a metaphorical description of the crime issue but also with pictures of different speakers that are associated with different levels of reliability, namely a *newscaster* and an aggressive looking, drunk person, who for the sake of simplicity will be called *hooligan* hereafter. The experiment was conducted using a 2 (metaphorical frame: *beast* or *virus*) x 2 (reliability: newscaster or hooligan) between-subjects design. The two options of the metaphorically framed description of crime, uttered by two different speakers, are given in the appendix. For each speaker, the text in the description of crime is slightly adjusted to match the speaker's jargon, as shown in the appendix.

After a welcoming page, participants were randomly presented with one of the four possible background scenarios describing crime in the City of Addison. While the description was still visible on the screen, participants were asked to propose solutions for the crime problem in a text box. Next, they were asked to rate the reliability of the text on a scale from 1 to 7. Finally, they were asked to indicate their political affiliation, age, education, gender, nationality, native language, and level of education. During the experiment, participants were not able to return to the pages they had already visited in the survey and update their responses. The code for the study can be found here: https://tinyurl.com/5etvk8h4.

As in the replication study, participants' solutions were treated as a set of suggestions. Each suggestion is classified as either *enforce* or *reform* by hand. As in the original study by Thibodeau & Boroditsky (2011), suggestions are categorized as *reform* if the proposed solution suggests investigating the underlying cause of the problem or suggests a particular social re-

form to treat or inoculate the community. In contrast, suggestions are categorized as *enforce* if the proposed solution focuses on the police force or other methods of law enforcement or modifying the criminal justice system. As pointed out by Thibodeau & Boroditsky (2015), suggestions of *neighborhood watches* cannot be clearly classified and are therefore excluded from the analysis. Solutions that contain the same number of suggestions for both reform and enforce are categorized as *both*. Finally, solutions are categorized as *neither* if the proposed solution lacked a suggestion that would fit in either of those categories. Participants' answers and their categorizations can be found here: https://tinyurl.com/2p9yv4h2. In contrast to the replication study, however, participants' responses are treated as *expected responses* if they suggested reform-oriented solutions in the *virus* condition or enforcement-oriented solutions in the *beast* condition.

As preregistered at https://osf.io/4fe9c, data were analyzed using a logistic regression model for the binary choice variable `speaker` as covariate (either *newscaster* or *hooligan*). The analysis uses the statistical programming language R. The code for the analysis can be found here: https://tinyurl.com/mrs64w68.

The preregistered study design is based on a pilot study that was completed by a total of 23 participants. As preregistered, we hypothesized that compared to the *hooligan* condition, we expect a higher proportion of expected suggestions (reform-oriented suggestions in *virus* frame and enforce-oriented suggestions in *beast* frame) in the *newscaster* condition. We judge there to be positive evidence in favor of the main hypothesis, if the posterior probability of the difference between expected responses in the reliable speaker condition and in the unreliable speaker condition (more expected responses in *newscaster* condition than in *hooligan* condition) being bigger than zero is at least 0.95.

4.2 Participants

As for the replication study, the data used in the follow-up study was taken form Prolific (www.prolific.co/) with an approval rate of 95% and only participants who self-identified as native English speakers based in the US were

recruited. Every participant was allowed to participate only once and was paid for their participation.

Due to technical issues, only 499 instead of the preregistered 500 participants completed the study, of whom 175 (35,1%) identified as females, 318 (63,7%) as males, 4 (0.8%) as other, and 2 (0.4%) preferred not to say. Their ages range from 18 to 82[2] with a mean of 41.5. A total of 150 participants (30.1%) completed high school as their highest level of education, while 7 (1.4%) participants did not graduate from high school. 244 participants (48.9%) reported college as their highest education, and 94 participants (18.8%) had a higher degree, while 4 preferred not to say. A total of 229 (45.9%) participants described themselves as Democrats, 107 (21.4%) as Republicans, while 157 (31.5%) identified as neither and 6 participants (1.2%) preferred not to say. The full data set can be found here https://tinyurl.com/5f6r59c5.

4.3 Findings

The aim of the follow-up study was to examine whether the effect of metaphors on reasoning is influenced by the (underlying) assumptions of the participants. That is, the use of a particular metaphor is not seen as a random decision but shows that the speaker has reflected on the issue and concluded that this metaphor best describes the problem, allowing participants to propose appropriate solutions.

Now, consider Figure 4, which shows participants' categorized responses to the crime issue. The answers of participants that were presented with the *newscaster* condition are presented on the top, while the answers of participants that were presented with the *hooligan* condition are displayed on the bottom.

---

[2] One participant reported an age of 1818. It can only be assumed that this data point was caused by a typing error. Therefore, the second highest value is given here as the highest age and the mean value of age is calculated without the outlier.
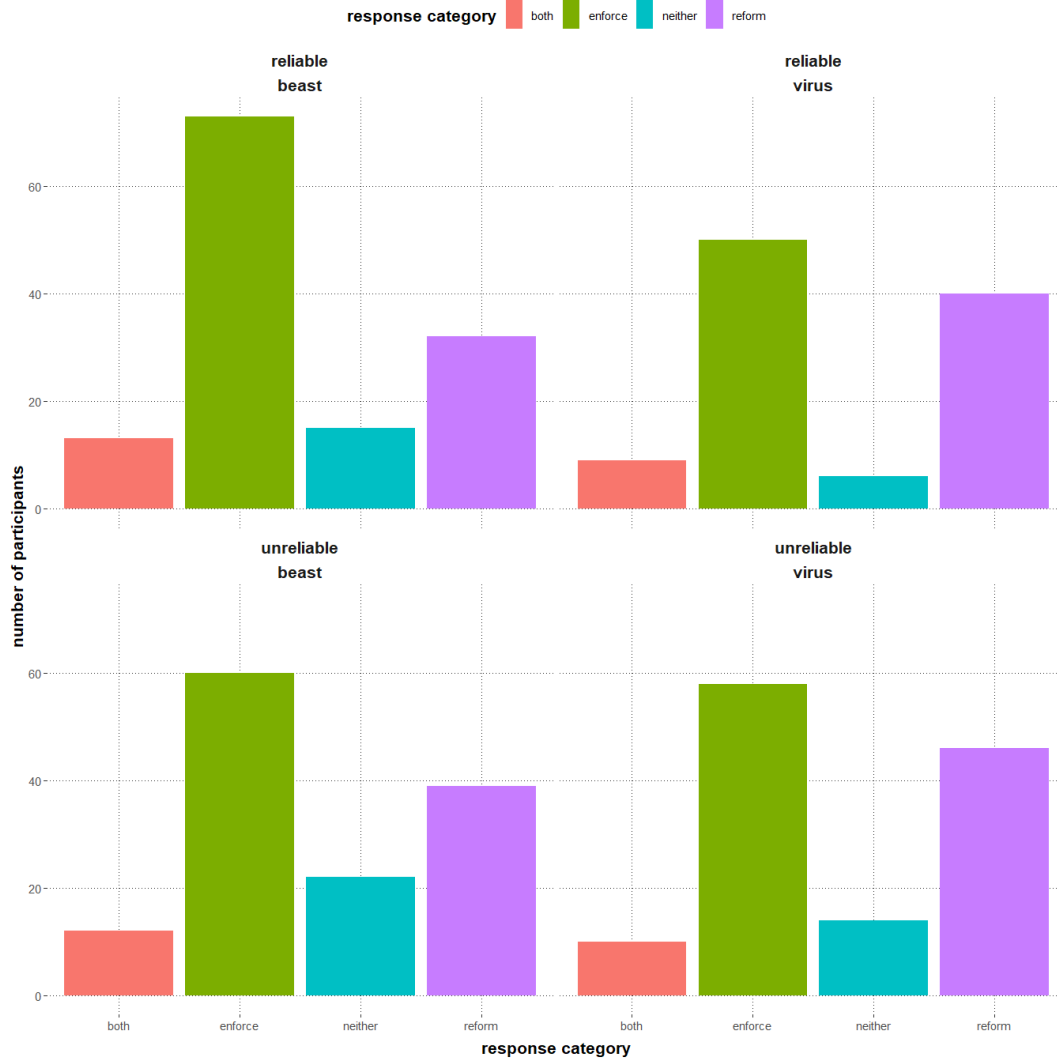
Figure 4: Participants' responses divided according to metaphorical frame and reliability of the speaker.

As preregistered, we hypothesized that compared to the *hooligan* condition, we expect a higher proportion of expected suggestions (reform-oriented suggestions in *virus* frame and enforce-oriented suggestions in *beast* frame) in the *newscaster* condition. We judge there to be positive evidence in favor of the main hypothesis, if the posterior probability of the difference between expected responses in the reliable speaker condition and in the unreliable speaker condition (more expected responses in *newscaster* condition than in *hooligan* condition) being bigger than zero is at least 0.95.

However, the results of the follow-up study fail to reach the required threshold, as the posterior probability is 0.90. Hence, the probability of the hypothesis, given our model and data, is less than 0.95 and therefore insufficient to support the hypothesis.

5. Discussion

In the following chapter, the key findings of the present study are presented, the main hypotheses are re-visited, and compared to previous studies in this field. Additionally, measurement errors, biases, unexpected findings and potential future research are discussed.

In the replication study, we find no evidence for the first hypothesis, that is, for the effect of metaphors on the choice of crime-reducing suggestions. Interestingly, as summarized in chapter 2, the only studies that could replicate said effect were conducted by at least one of the two co-authors of the original study by Thibodeau & Boroditsky (2011), while other replication studies found no such effect. The only exception is Christmann & Göhring (2016). However, while they conclude that their results replicate said effect, the coding of answers differed, as described in more detail in chapter 2. They admit that "conducting the analysis in full accordance with the original procedure[,] this difference was not statistically meaningful" (ibid.).

In contrary, our results provide evidence for the second hypothesis that Republicans are more likely to propose enforcement-oriented suggestions, as claimed in the original study. This effect could be replicated in previous studies as well, including Reijnierse et al. (2015), who did not find the metaphor-framing effect but the effect of political affiliation as well.

Regarding our last hypnosis, we used two speakers that are associated with different levels of reliability to examine the impact of the speaker's reliability on the metaphor-framing effect. As proposed by Thibodeau & Boroditsky (2011), even subtle metaphors can influence how we reason about a particular issue. However, little research has been conducted regarding the question how exactly this effect is established. On the one hand, one could say that using a certain metaphor causes the hearer to associate the issue with the source domain of said metaphor. That is, the activation of representations for solving an infection with a virus or an attack by a beast are transferred to the crime problem. According to this theory, it is the metaphoric word itself that influences how the hearer reasons about the issue

by activation certain associations. Therefore, the context in which the metaphor was uttered is irrelevant to how influential the metaphor is but using the metaphor several times increases the influence that the metaphor exerts.

On the other hand, on could say that using a certain metaphor does not necessarily cause the hearer to associate the problem with a certain source domain. Instead, the influence of the metaphor varies with the context. That is, if the speaker seems reliable, the hearer concludes that the metaphor best describes the problem and thus, allowing the hearer to propose appropriate solutions that are in line with the metaphor. In contrast, if the hearer does not seem reliable, the hearer does not believe that metaphor suits the problem and therefore, rejects the associations induced by the metaphorical framing of the problem and is less likely to propose metaphor-consistent solutions. In this theory, hearing the metaphorically framed description of crime by a newscaster would lead hearers to prefer metaphor-consistent solutions, while this effect is less prominent in the condition with the hooligan as speaker.

Our results revealed that the difference between the two speakers (0.90) closely dismisses the threshold of 0.95. That raises the question if it is, indeed, only the metaphoric word itself that activates certain association, as proposed in the theory above, or if external factors might have caused the results to not reach the required threshold. In the following, therefore, several biases and measurement errors that might have occurred in the present are discussed.

First, it cannot be ruled out that various news stories, political discourses, and other experiences related to crime that the subjects experienced shortly before the study have affected on their responses.

Second, the present study uses internet recruitment, which is associated with certain biases. For instance, only people with access to internet, the required time to fill in the survey, and interest in participation are recruited. As a result, the sample might under-represent certain groups of people. However, conducting the study online perhaps reduced the risk of

participants changing their responses due to being observed by the re-searchers. This is less likely to be the case in the present study, since the answers were collected anonymously online.

Next, as pointed out by Thibodeau & Boroditsky (2015) cultural shifts can alter "how people conceptualize aspects of crime" (p. 11). They found that this cultural shift must have occurred in the past six years (ibid.). Since the present study is conducted more than six years after their latest replication study and twelve years after the original study, it is likely that another cultural shift has occurred since that might have influenced how crime is perceived nowadays and explains why the same results as in the original study could not be replicated.

Moreover, in the replication study as well as in the follow-up study, we found an overall tendency for enforcement-oriented suggestions, as shown in Figure 1 in chapter 3.3 for the replication study and below for the follow-up study.
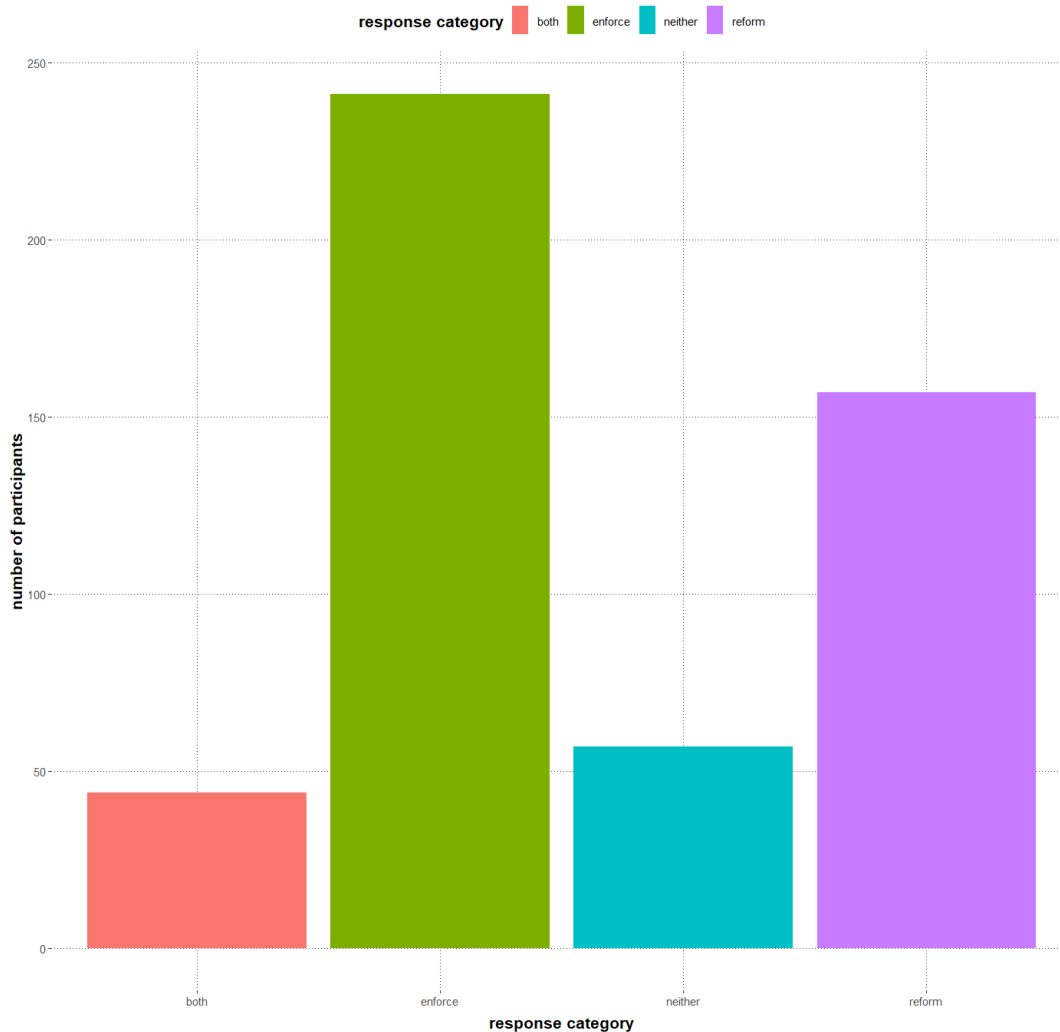
Figure 5: Overall preference of enforcement-oriented suggestions in the follow-up study.

Steen, Reijnierse & Burgers (2014) point out that such a pattern "could lie in the construction of the particular stimulus text used as well as in the dependent variable" (p. 21). The set-up of the study, therefore, might have influenced the data, not allowing the results to reach the required threshold.

Furthermore, the distinction between *reform* and *enforce* is unclear in some cases, insofar as the distinction proposed in the original study only differentiates between measurement to investigate the root causes for crime and measurement that punish criminals does not allow for measurements that do not prevent crime but hinder criminals from committing crime, e. g. security measures such as increased lighting in public areas and improved security measures at private houses. This raises the question whether deterring crime counts as preventative strategy or as enforcement strategy. It

is unclear, whether the results were the same, if those responses with the measurements in question were categorized differently.

Similarly, according to the original study, investigating the root cause, should be seen as reform-oriented suggestions (cf. Thibodeau & Boroditsky, 2011). However, a lot of participants' responses did not suggest scrutinize the root of the problem but what types of crimes are committed. It can only be assumed that investigating the type of crimes implies to, as a next step, also investigate the root of those crimes but since this is only an assumption, those suggestions cannot be clearly categorized either.

In addition, Steen, Reijnierse & Burgers (2014) point out that according to the metaphor processing termination hypothesis, the influence of metaphors depends on the complexity of the described problem. That is, the effect of metaphors is reduced to cases, in which the metaphor is needed to understand the problem. Since crime is a frequently discussed topic that assumingly all participants have reflected upon prior to the study, it might be that this is why we found no evidence supporting the hypothesis. However, if that is the case, then the question arises why such an effect was found in the original study by Thibodeau & Boroditsky, since their experiment was regarding crime as well.

Besides, in the follow-up study, the manipulation of vignettes is not reduced to using different pictures of speakers to portrait different levels of reliability. Instead, the text in the description is slightly adjusted and the different pictures could activate certain associations that cannot be controlled for the researchers. That is, since the unreliable speaker is depicted by an obviously drunk, aggressive looking person, this might have triggered participants to propose other suggestions than would have been the case if the unreliable speaker were illustrated differently. Similarly, for the reliable speaker, we chose a newscaster. However, not all participants might consider newscaster the most reliable source for information of crime. For example, consider one participant's response shown in (1).

(1)   Get a new reporter! Too many opinion words were used in the report to scare people

Therefore, this might have influenced the results as well.

As a last point, while the focus of the present study lays on the four hypotheses, the results indicated certain unexpected findings that are addressed in the following. Many participants suggested increasing number of police officers as a crime-reducing strategy. According to Thibodeau & Boroditsky (2011), suggestions focused on police force should be classified as enforcement-oriented. However, it becomes apparent in participants' responses that police patrols are seen as a preventative measure instead. That is, increasing the police presence in the streets of Addison, would deter crime by scaring of criminals. For instance, consider two participants' responses shown in (2).

(2) They need to put more police officers in the street to prevent crime[.]

(3) Addison needs to hire more police to patrol the streets and answer calls, especially at night. If criminals know that there are [sic] police near by [sic] and they will be caught, they will be less likely to commit a crime.

In conclusion, it can be said that there are two forms of prevention. First, prevention in the sense of reform-oriented suggestions includes measures to avoid crime at the root. Second, prevention in the sense that is mentioned by some participants suggesting police patrols, includes measures that do not reduce crime at the root but deter crime because of fear or because of possible trouble or difficulty. Further research is required that either investigates a metaphoric frame that leaves less room for interpretation or explores all different categories of responses.

6. Conclusions

Not only the way in which a problem is described, but also the contextual circumstances of the description can influence how we reason about said problem. In two studies, we investigated the effect of metaphors on reasoning. The first study aimed to explore the role of metaphors and political affiliations, while the second study analyzed how the reliability of the speaker influences the effect of the metaphor.

In the first study, we found no support for the effect shown by Thibodeau & Boroditsky (2011) that the metaphorically framed description

influences the sort of suggestion offered by participants. However, we found that participants who self-identify as Republicans are more likely to propose enforcement-oriented suggestions than other participants, which is consistent with the results presented by Thibodeau & Boroditsky (2011). Additionally, our results indicate that the shorter vignette, which is used in the follow-up study, does not correlate with the perceived reliability of the text. Moreover, the results of the second study indicate that (underlying) assumptions of the participants do not influence the effect of metaphors on reasoning.

- Make an argument in support of your overall conclusion
  - What is my overall conclusion?
    - Our results indicate that the context does not seem to affect the influence of the metaphor.
    - This leads to the assumption that metaphoric words themselves activate certain associations, as proposed in the theory above.
    - Further research is required to explore how metaphors work in more depth.

7. Bibliography

Bougher, L. D. (2012). The Case for Metaphor in Political Reasoning and Cognition: Metaphor in Political Reasoning and Cognition. *Political Psychology*, *33*(1), 145–163. https://doi.org/10.1111/j.1467-9221.2011.00865.x

van den Broek, B. (2005). *Treating the virus and capturing the beast: The effect of verbal and visual metaphor on reasoning about social problems.* Master's thesis, Tilburg University.

Christmann, U., & Göhring, A.-L. (2016). A German-language replication study analysing the role of figurative speech in reasoning. *Scientific Data*, *3*(1), 160098. https://doi.org/10.1038/sdata.2016.98

Fausey, C. M. (2010). Constructing agency: The role of language. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00162

Flusberg, S. J., Lauria, M., Balko, S., & Thibodeau, P. H. (2020). Effects of Communication Modality and Speaker Identity on Metaphor Framing. *Metaphor and Symbol*, *35*(2), 136–152. https://doi.org/10.1080/10926488.2020.1767336

Flusberg, S. J., Matlock, T., & Thibodeau, P. H. (2018). War metaphors in public discourse. *Metaphor and Symbol*, *33*(1), 1–18. https://doi.org/10.1080/10926488.2018.1407992

Gibbs, R. W. (o. J.-a). *The Cambridge Handbook of Metaphor and Thought*.

Gibbs, R. W. (2015). Do pragmatic signals affect conventional metaphor understanding? A failed test of deliberate metaphor theory. *Journal of Pragmatics*, *90*, 77–87. https://doi.org/10.1016/j.pragma.2015.05.021

Heyvaert, P., Randour, F., Dodeigne, J., Perrez, J., & Reuchamps, M. (2020). Metaphors in political communication: A case study of the use of deliberate metaphors in non-institutional political interviews. *Journal of Language and Politics*, *19*(2), 201–225. https://doi.org/10.1075/jlp.17057.hey

Johannesson, N.-L. (o. J.). *When is metaphor deliberate*.

Knowles, M., & Moon, R. (2004). *Introducing Metaphor* (0 Aufl.). Routledge. https://doi.org/10.4324/9780203642368

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3

Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O. (2013). Metaphor Identification in Large Texts Corpora. *PLoS ONE*, *8*(4), e62343. https://doi.org/10.1371/journal.pone.0062343

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2015). How viruses and beasts affect our opinions (or not): The role of extendedness in metaphorical framing. *Metaphor and the Social World*, *5*(2), 245–263. https://doi.org/10.1075/msw.5.2.04rei

Steen, G. (2008). The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor. *Metaphor and Symbol*, *23*(4), 213–241. https://doi.org/10.1080/10926480802426753

Steen, G. (2009). Deliberate Metaphor Affords Conscious Metaphorical Cognition. *Cognitive Semiotics*, *5*(1–2), 179–197. https://doi.org/doi:10.1515/cogsem.2013.5.12.179

Steen, G. (2017). Deliberate Metaphor Theory: Basic assumptions, main tenets, urgent issues. *Intercultural Pragmatics*, *14*(1), 1–24. https://doi.org/10.1515/ip-2017-0001

Steen, G. J. (2011). The contemporary theory of metaphor—Now new and improved! *Review of Cognitive Linguistics*, *9*(1), 26–64. https://doi.org/10.1075/rcl.9.1.03ste

Steen, G. J., Reijnierse, W. G., & Burgers, C. (2014). When Do Natural Language Metaphors Influence Reasoning? A Follow-Up Study to Thibodeau and Boroditsky (2013). *PLoS ONE*, *9*(12), e113536. https://doi.org/10.1371/journal.pone.0113536

Thibodeau, P. H. (2016). Extended Metaphors are the Home Runs of Persuasion: Don't Fumble the Phrase. *Metaphor and Symbol, 31*(2), 53–72. https://doi.org/10.1080/10926488.2016.1150756

Thibodeau, P. H., & Boroditsky, L. (2013). Natural Language Metaphors Covertly Influence Reasoning. *PLoS ONE, 8*(1), e52961. https://doi.org/10.1371/journal.pone.0052961

Thibodeau, P. H., & Boroditsky, L. (2015a). Measuring Effects of Metaphor in a Dynamic Opinion Landscape. *PLOS ONE, 10*(7), e0133939. https://doi.org/10.1371/journal.pone.0133939

Thibodeau, P. H., & Boroditsky, L. (2015b). Measuring Effects of Metaphor in a Dynamic Opinion Landscape. *PLOS ONE, 10*(7), e0133939. https://doi.org/10.1371/journal.pone.0133939

Thibodeau, P. H., Crow, L., & Flusberg, S. J. (2017). The metaphor police: A case study of the role of metaphor in explanation. *Psychonomic Bulletin & Review*, *24*(5), 1375–1386. https://doi.org/10.3758/s13423-016-1192-5

Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor Aptness and Conventionality: A Processing Fluency Account. *Metaphor and Symbol*, *26*(3), 206–226. https://doi.org/10.1080/10926488.2011.583196

Thibodeau, P. H., & Flusberg, S. J. (In press). *Metaphor and Elaboration in Context*. John Benjamins Publishing. https://osf.io/c93nf/download

Thibodeau, P. H., Flusberg, S. J., Glick, J. J., & Sternberg, D. A. (2013a). An emergent approach to analogical inference. *Connection Science*, *25*(1), 27–53. https://doi.org/10.1080/09540091.2013.821458

Thibodeau, P. H., Flusberg, S. J., Glick, J. J., & Sternberg, D. A. (2013b). An emergent approach to analogical inference. *Connection Science*, *25*(1), 27–53. https://doi.org/10.1080/09540091.2013.821458

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How Linguistic Metaphor Scaffolds Reasoning. *Trends in Cognitive Sciences*, *21*(11), 852–863. https://doi.org/10.1016/j.tics.2017.07.001

Thibodeau, P. H., Matlock, T., & Flusberg, S. J. (2019). The role of metaphor in communication and thought. *Language and Linguistics Compass*, *13*(5), e12327. https://doi.org/10.1111/lnc3.12327

Thibodeau, P., Peebles, M. M., Grodner, D. J., & Durgin, F. H. (2015). The Wished-For Always Wins Until the Winner Was Inevitable All Along: Motivated Reasoning and Belief Bias Regulate Emotion During Elections: Elections Bias Reasoning. *Political Psychology*, *36*(4), 431–448. https://doi.org/10.1111/pops.12100
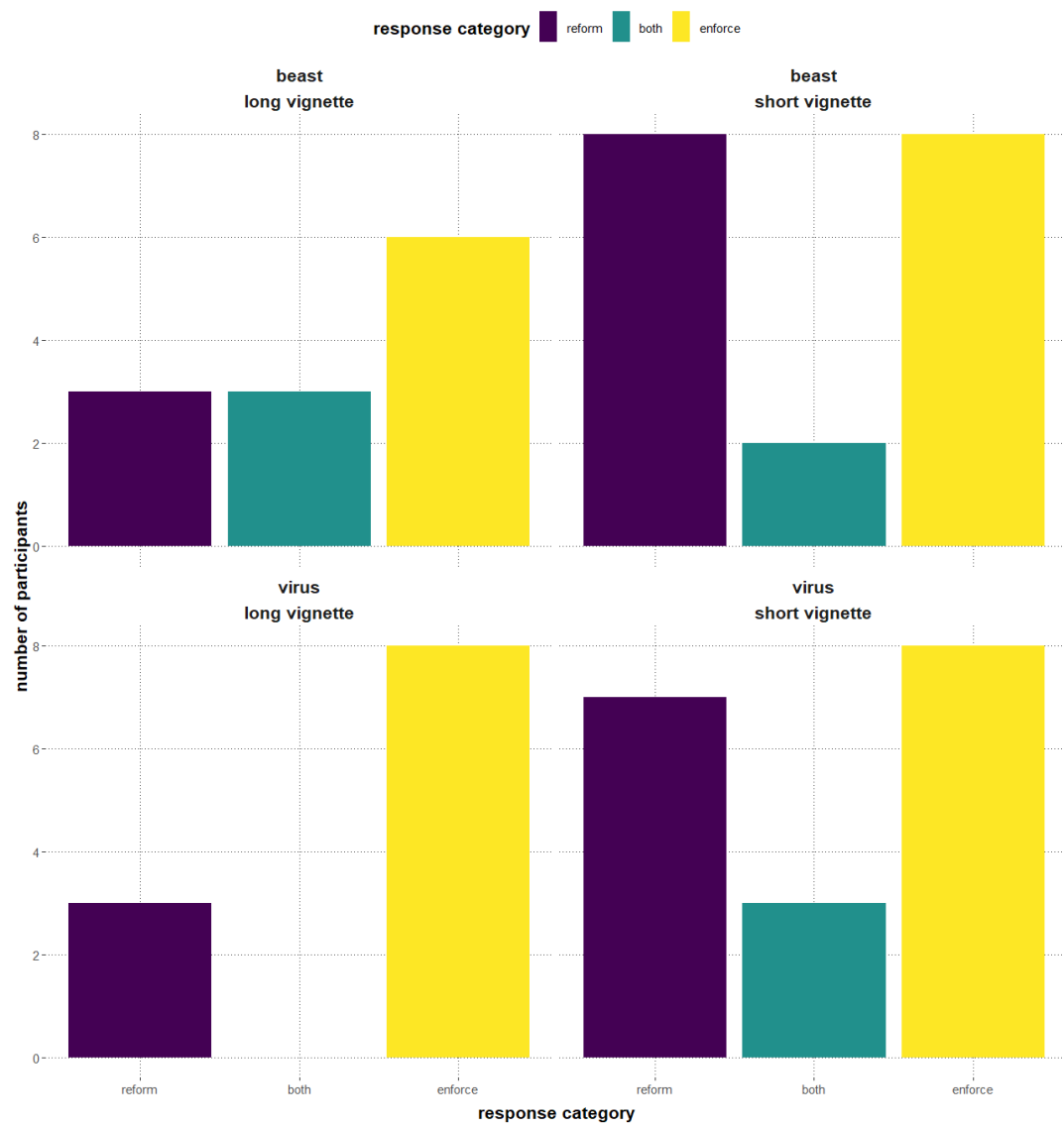
8. Appendix

8.1 Pilot study: Findings

Figure 6: Perceived reliability of vignette in pilot study of replication study
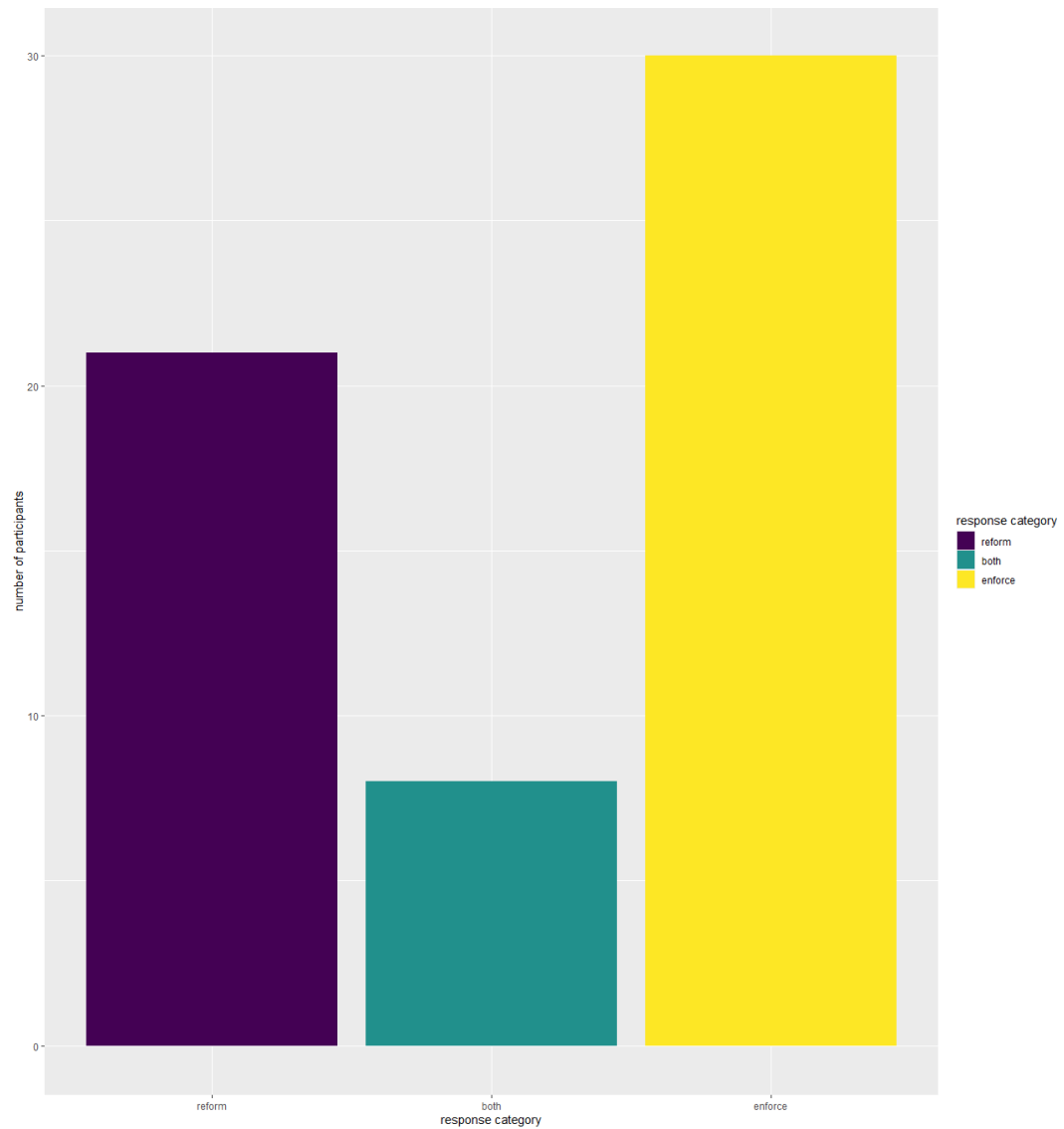
Figure 7: Participants in pilot study of replication study favor enforcement-oriented measurements overall.

Figure 8: Participants' categorized responses in pilot study of replication study.

8.2 Materials of replication study

Virus, long vignette:

Crime is a virus infecting on the city of Addison. The crime rate in the once peaceful city has steadily increased over the past three years. In fact, these days it seems that crime is plaguing every neighborhood. In 2004, 46,177 crimes were reported compared to more than 55,000 reported in 2007. The rise in violent crime is particularly alarming. In 2004, there were 330 murders in the city, in 2007, there were over 500.

Virus, short vignette:

Crime is a virus infecting on the city of Addison. The crime rate has steadily increased over the past three years. In fact, these days it seems that crime is plaguing every neighborhood. The rise in violent crime is particularly alarming.
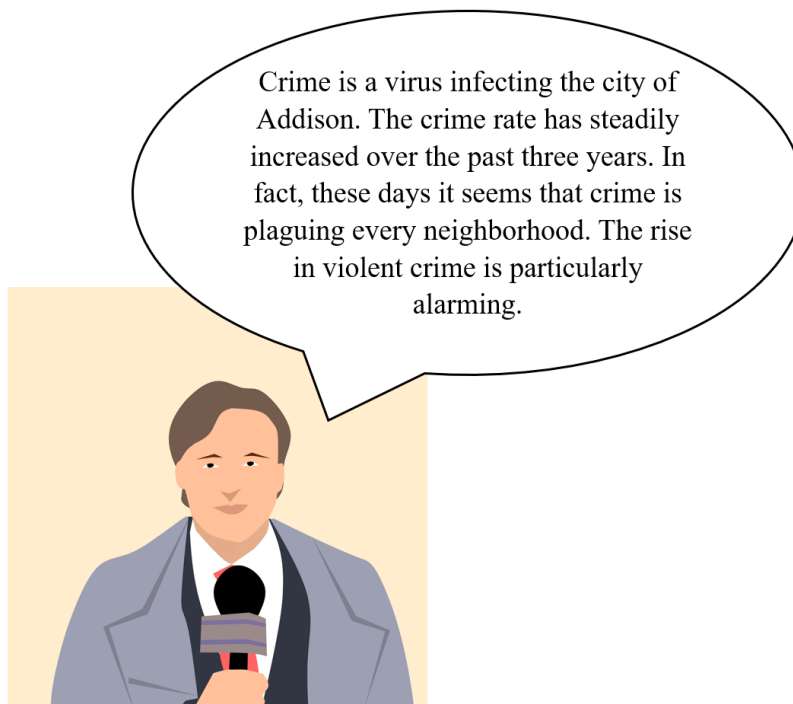
Beast, long vignette:

Crime is a wild beast preying on the city of Addison. The crime rate in the once peaceful city has steadily increased over the past three years. In fact, these days it seems that crime is lurking in every neighborhood. In 2004, 46,177 crimes were reported compared to more than 55,000 reported in 2007. The rise in violent crime is particularly alarming. In 2004, there were 330 murders in the city, in 2007, there were over 500.

Beast, short vignette:

Crime is a wild beast preying on the city of Addison. The crime rate has steadily increased over the last three years. In fact, these days it seems that crime is lurking in every neighborhood. The rise in violent crime is particularly alarming.

8.3 Materials of follow-up study

Virus, reliable speaker:

Crime is a virus infecting the city of Addison. The crime rate has steadily increased over the past three years. In fact, these days it seems that crime is plaguing every neighborhood. The rise in violent crime is particularly alarming.

Virus, unreliable speaker:

Crime is a virus infecting the city of Addison. The crime rate has steadily increased over the past three years. Not gonna lie, now it seems that crime is plaguing every damn neighborhood. The rise in violent crime is fucking crazy.
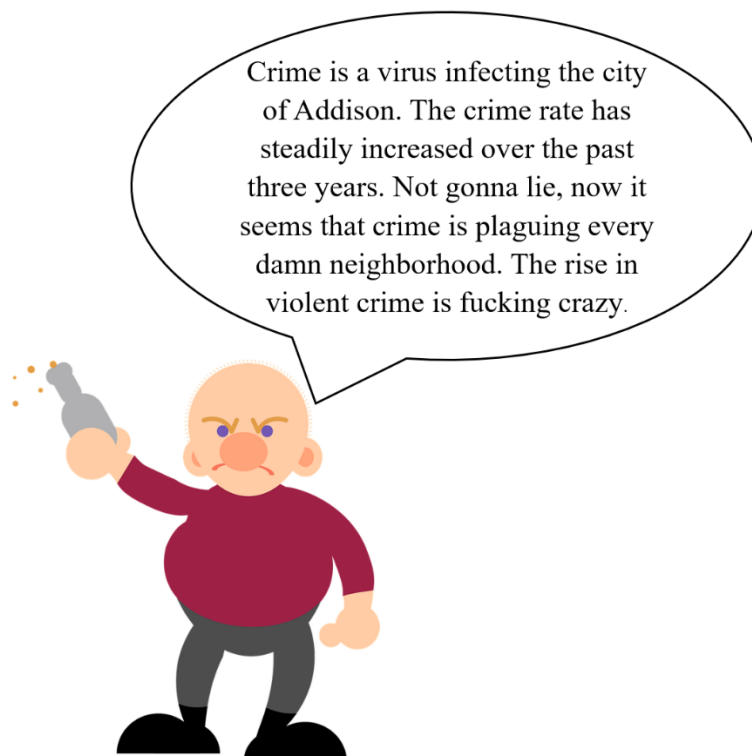
Beast, reliable speaker:

Crime is a wild beast preying on the city of Addison. The crime rate has steadily increased over the last three years. In fact, these days it seems that crime is lurking in every neighborhood. The rise in violent crime is particularly alarming.
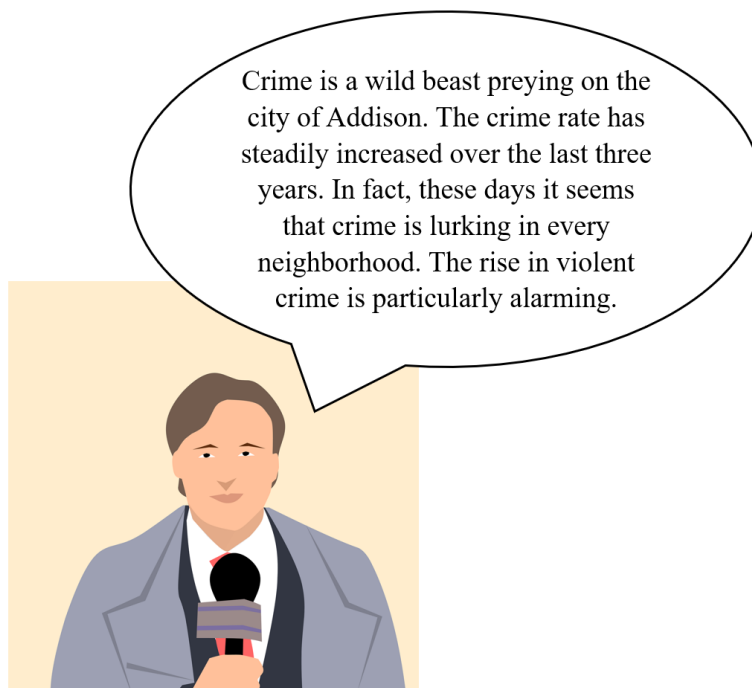
Beast, unreliable speaker:

Crime is a wild beast preying on the city of Addison. The crime rate has steadily increased over the last three years. Not gonna lie, now it seems that crime is lurking in every damn neighborhood. The rise in violent crime is fucking crazy.

Declaration of Authorship

I hereby confirm that this paper and the work presented in it is entirely my own. Where I have consulted the work of others this is always clearly stated. All statements taken literally from other writings or referred to by analogy are marked and the source is always given. This paper has not yet been submitted to another examination office, either in the same or similar form.

Tübingen, April 17th, 2023

Miriam Schiele

Miriam Schiele