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

While the persuasive nature of metaphors is well-known, previous research has focused on the influence of the metaphoric words, neglecting the impact of context and the underlying assumptions of the participants. In two experiments (N = ?), we examined the influence of metaphoric framing and speakers on reasoning about a socio-political problem, namely crime. In the first experiment, participants read a short, metaphorically framed description of a city's crime problem, in which crime is either referred to as *beast* or *virus*. In the second experiment, the description was illustrated as a statement by speakers that are associated with different levels of reliability about said problem, maintaining the metaphorical framing. In an open-end format, participants were asked to propose crime-reducing suggestions.

Our results indicate that the efficacy of metaphors is influenced by a number of factors and that the metaphoric effect is smaller than suggested in previous studies. For instance, the political affiliation plays a larger role than the metaphoric frame. These results provide new insights into the efficacy and application of the impact of metaphor.

while several previous studies explored the metaphor framing effect on reasoning, the question of whether reliability of the speaker influences hearer's reasoning about an issue has not been investigated in the context of metaphors so far.

1. Introduction

Metaphors can be found in everyday speech. Their persuasive nature has been object of studies since the last 1970s (cf. Gibbs, 2008). 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. Then, I will present studies that could not replicate the findings and discuss their results. Finally, studies investigating whether speaker reliability affects people's judgments are reviewed.

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 United States 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).

Regarding speaker reliability, the following studies investigated the effects that speakers' trustworthiness can have on participants' judgements. However, neither of the following two studies have been peer-reviewed. They are mentioned here only as an indicator for how speaker reliability may affect decision making. See chapter 5 for an in-depth discussion.

Neal (2015) asked participants to read a scenario describing rape. The descriptions solely differed in the target word (euphemism or no euphemism), the placement of the target word within the description, and in the speaker (victim or lawyer). While the results showed no significance for the effect of the language use, the difference in speaker that were associated

with different levels of reliability significantly influenced participants' judgements. Hence, while the studies mentioned above focused on the influence of metaphor on decision making, Neal (2015) proves that it is not the language, but rather the context that affects people's judgement.

Lastly, Teixeira (2022), who explored influencer marketing, found that the trustworthiness of influencers defined as "the degree of confidence in the communicator's intent to communicate the assertions he considers most valid" (Hovland et al., 1953, p. 21, cited by Teixeira, 2022, p. 23) has the most powerful effect on purchase intention, that is, how likely the customer is to buy the product. Even though this study does not examine participants' judgement on crime, it highlights the importance of speakers' credibility.

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, sociodemographic 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.

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. 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 preregistered study design is based on a pilot study that was completed by a total of 60 participants. The code for the experiment, the code for the data analysis, participants' answers, their categorization, the findings of the pilot study as well as of the replication study can be found here: https://bit.ly/40MWhJe.

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 United States 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.

---

[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.

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.
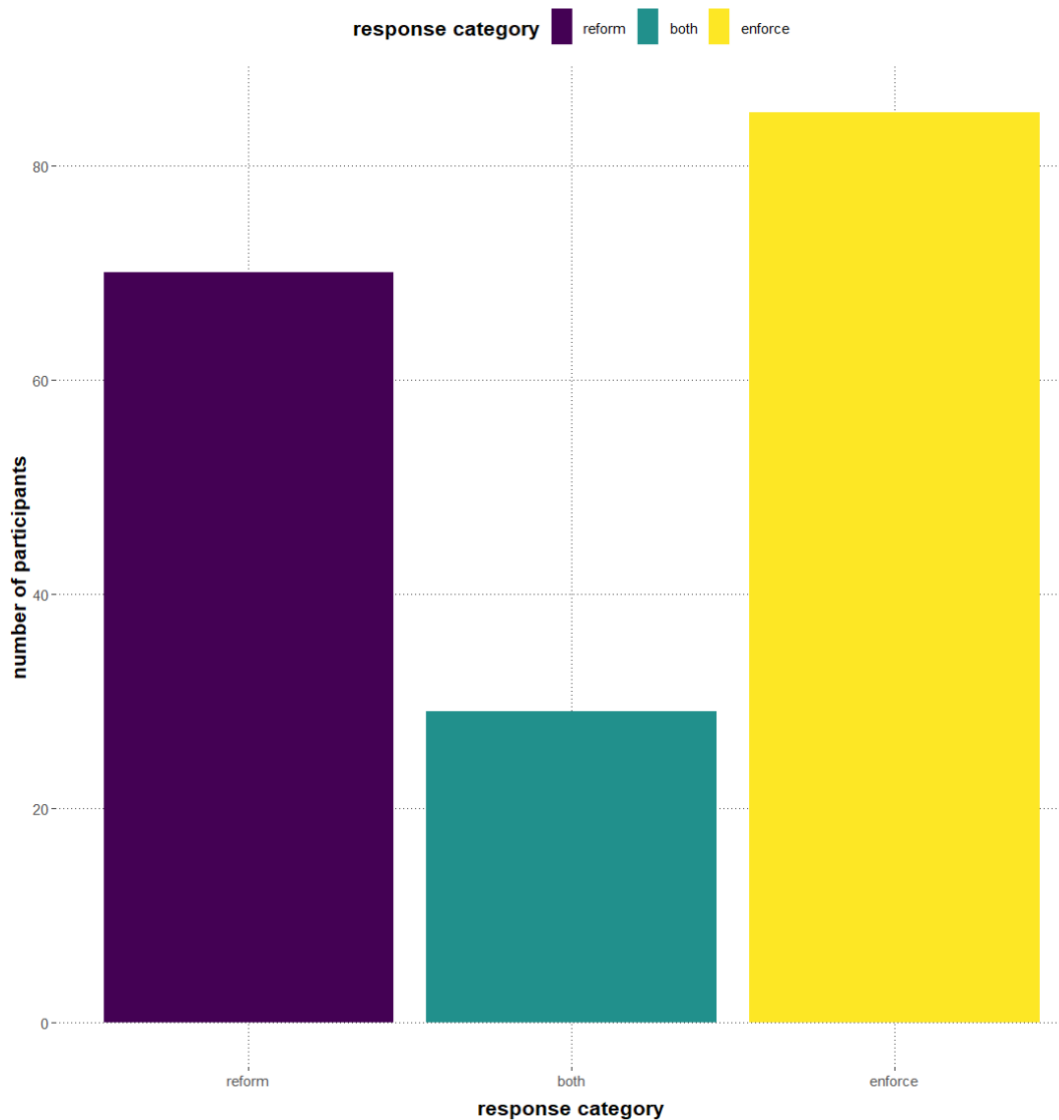
Figure 1: Categorizing participants' responses reveals that enforcement-oriented solutions are preferred overall in the replication study.

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.
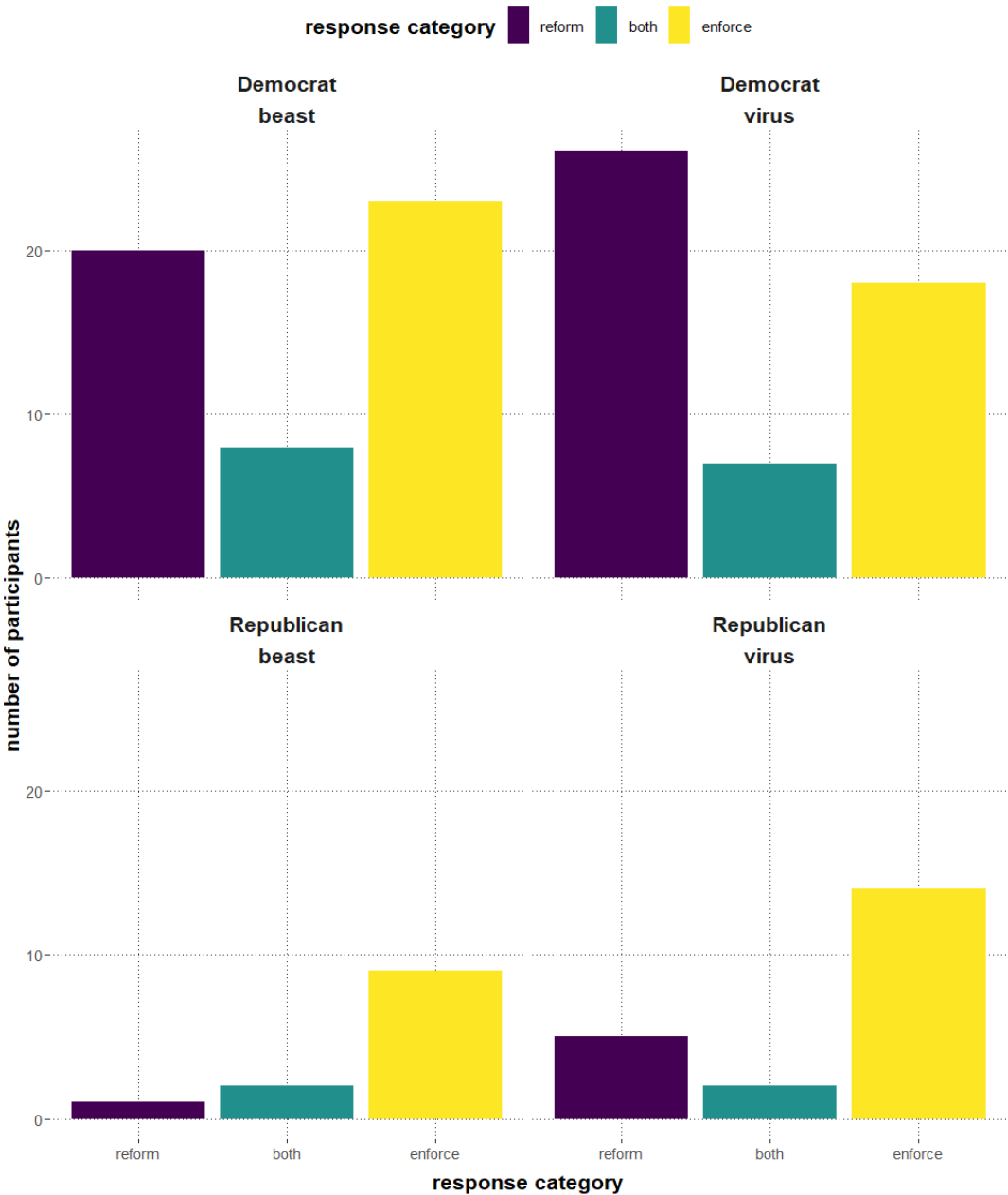
Figure 2: Democrats prefer enforcement-oriented solutions overall, whereas Republicans are more likely to suggest reform-oriented suggestions, independently of the metaphoric frame.

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 ($r = 4.83$ in the long vignette condition and $r = 5.12$ in the short vignette condition), although the results reveal that the variation is rather high ($\sigma = 1.45$ and $\sigma = 1.33$, respectively).
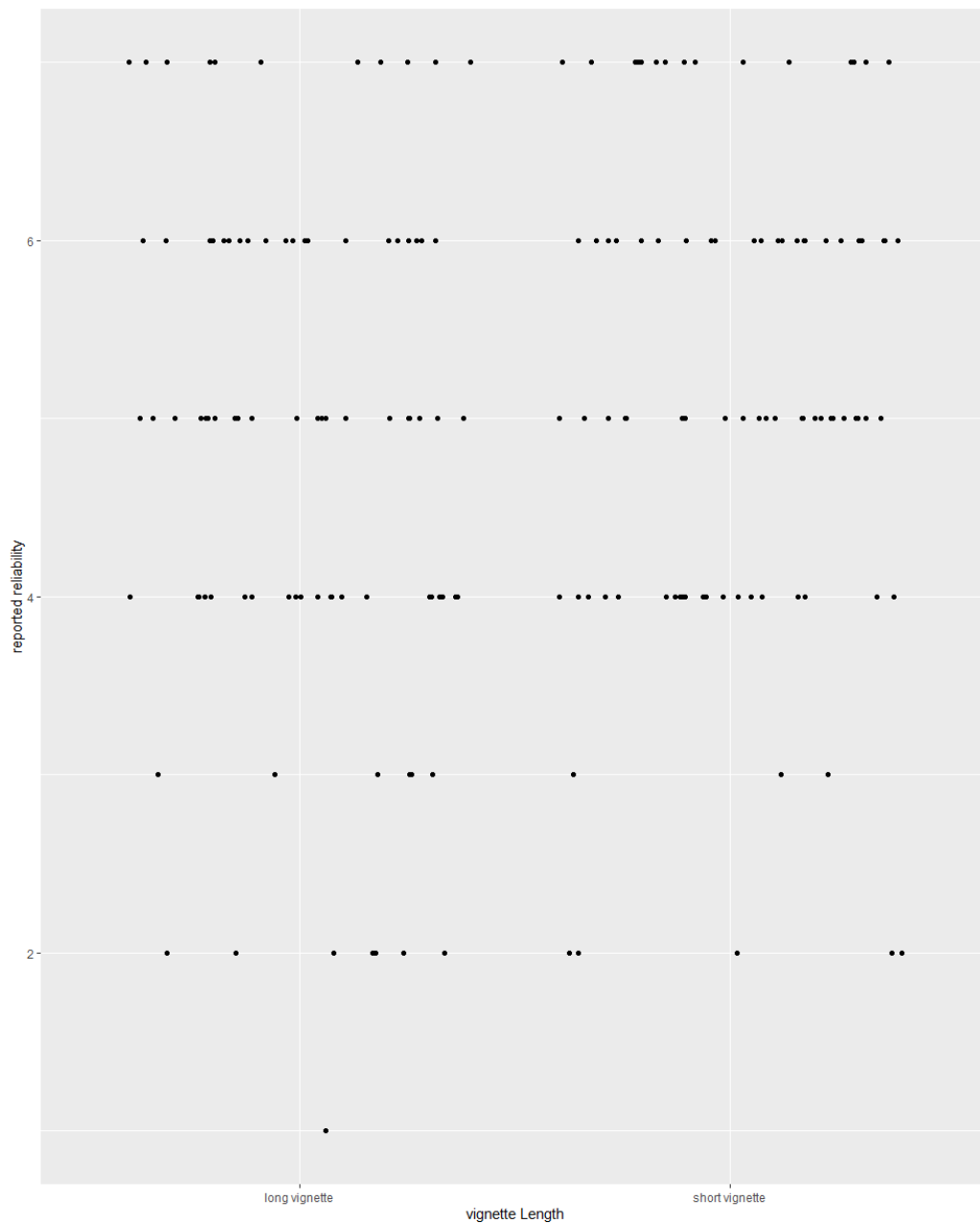
Figure 3: Plotting reliability against vignette length reveals hardly any influence of vignette length on perceived reliability in replication study.

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.

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 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 that would fit in either of those categories. 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 experiment, the code for the data analysis, participants' answers, their categorization, the findings of the pilot study as well as of the replication study can be found here: https://bit.ly/3zBDTY2.

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 United States 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.
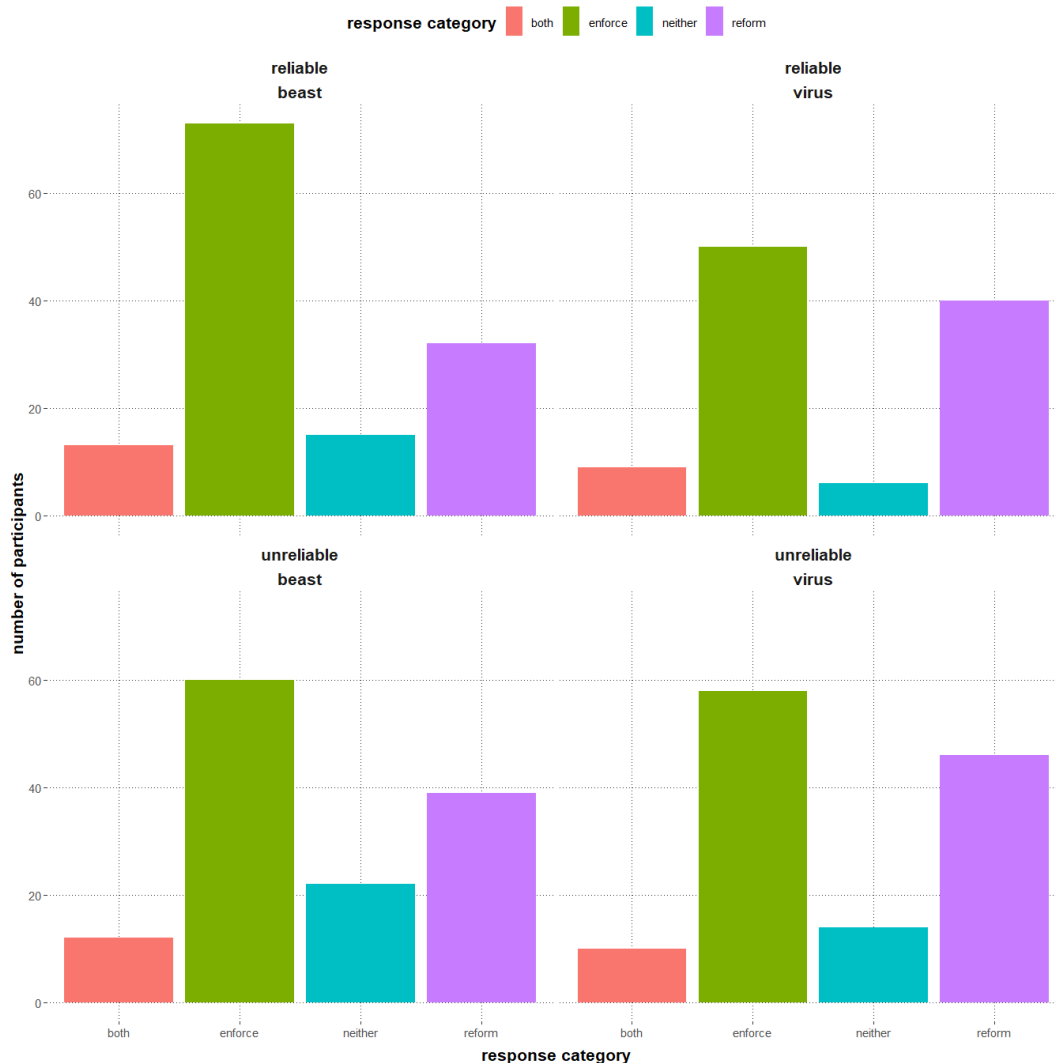
Figure 4: Metaphorical frame and reliability of the speaker do not seem to heavily influence participants' responses.

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. The mild trend (0.90) indicates that further

studies are required that compare the present results to other regression models and conduct causal analyses explore the impact of contextual factors in more depth. However, as preregistered, such methodologies and further studies are not part of the present study.

5. Discussion

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

5.1 Re-visiting the hypotheses

In what follows, the results of the two experiments are reflected regarding the hypotheses formulated in chapter 3 and 4 and regarding previous studies. 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 were able to replicate this 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 (cf. Steen, Reijnierse & Burgers, 2014, Reijnierse et al., 2015, van den Broek, 2015). The only exceptions are Flusberg et al. (2020) and Christmann & Göhring (2016). However, the former study did not use open-end format to capture participants' crime-reducing suggestions, but only allowing participants to choose between only two response options. As a result, the measure of the dependent variable differs greatly from the original study by neglecting the option for responses that do not fit into either category or that would include both enforcement-oriented and reform-oriented suggestions. Similarly, in the latter study, the measure differed from the original study, as described in more detail in chapter 2. Christmann & Göhring (2016) conclude that their results replicate the aforementioned effect. However, they concede that when "conducting the analysis in full accordance with the original procedure[,] this difference [between metaphor-consistent and metaphor-inconsistent suggestions] was not statistically meaningful" (ibid., p. 5).

Regarding our second hypothesis, our results provide evidence that Republicans are more likely to formulate enforcement-oriented proposals, as claimed in the original study. This effect could be replicated in previous studies, including Reijnierse et al. (2015). In their results, they were only able to detect the political affiliation effect, but not the metaphor framing effect. This is in sharp contrast to Thibodeau & Boroditsky's (2011) findings that the metaphor framing effect is larger than differences between Democrats and Republicans. Instead, it appears that while political affiliation plays a large role in decision making on the given socio-political issue, the metaphor framing may be modest and easily influenced by confounding factors. The results of the present study suggest that different interpretations of the metaphor, prior knowledge about the topic, and the complexity of the issue may play a role. However, based on our results, it can be assumed that other factors (some of which are yet to be identified) are likely to have an impact on the metaphor framing effect. These factors in questions are discussed in chapter 5.2.

To test our last hypnosis, we used two speakers associated with different levels of reliability to examine the impact of the speaker's reliability on the metaphor framing effect. As suggested by Thibodeau & Boroditsky (2011), even subtle metaphors can influence how we reason about a particular topic. However, the question of how exactly this effect arises and which factors play a role in it is still unanswered. On the one hand, Thibodeau, Hendricks & Boroditsky (2016) suggest that using a certain metaphor causes the hearer to associate the topic with the metaphor's source domain. That is, representations of solving an infection caused by a virus or an attack by a beast are activated and transferred to the crime problem. According to this theory, it is the metaphorical word itself that influences how the hearer reasons about the issue by activating certain associations. Therefore, the context in which the metaphor is uttered is irrelevant to the influence of the metaphor, but multiple use of the metaphor increases the influence that the metaphor exerts.

On the other hand, it could be argued for that the use a particular metaphor does not necessarily lead the hearer to associate the problem with a certain source domain. Rather, the influence of the metaphor varies with the

context. According to the Cooperative Principle, "[p]articipants expect that each will make a conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange." (Grice, 1975, p. 45). Therefore, if the speaker appears trustworthy, the hearer concludes that the speaker purposely chose this metaphor because it best describes the problem, ensuring a smooth conversation in correspondence with the Cooperative Principle. Thus, since the metaphor fits to illustrate the problem, it allows the hearer to propose appropriate solutions that are in line with the metaphor. If, in contrast, the speaker appears to be untrustworthy, the hearer does not believe that the metaphor suits the problem and therefore, rejects the associations evoked by the metaphorical framing of the problem, making the hearer less likely to suggest solutions that match the metaphor. According to this theory, listening to a newscaster's metaphorically framed description of crime, would lead hearers to prefer metaphor-consistent solutions since the speaker seems like a trustworthy interlocutor who obeys to the Cooperative Principle, whereas the metaphor framing effect is less pronounced in the condition with the drunk, aggressive looking speaker who presumably is regarded as less trustworthy and less likely to obey the Cooperative Principle.

Our results reveal that the difference between the two speakers (0.90) closely dismisses the threshold of 0.95. Therefore, speaker reliability does not seem to heavily influence participants' judgements. This stand in contrast with the findings by Neal (2015) and Teixeira (2022), who both supported the hypothesis that speaker reliability significantly effects participants' judgements. However, since neither of the two studies was peer-reviewed, further in-depth, high-quality research in this field is required. Additionally, our results raise the question of whether it is indeed only the metaphorical word itself that triggers certain associations, as suggested in the first theory above, or whether external factors may have impacted participants' decision making more than the reliability of the speakers.

5.2 Confounding factors, biases, and other influencing factors

As mentioned above, the metaphor framing effect as well as the speaker reliability effect seem to be modest and can be easily influenced by confounding factors. Therefore, several external factors, biases, and measure errors that may have occurred in the present study are discussed below. First, as Thibodeau & Boroditsky (2015) point out, cultural shifts can alter "how people conceptualize aspects of crime" (p. 11). They noted 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 most recent replication study and twelve years after the original study, it is likely that different cultural shift has occurred since then that may have influenced the contemporary perceptions of crime and explains why the same results as in the original study could not be replicated.

Second, the influence of metaphors is not only affected by cultural shifts that occur over time within the same culture, but as Eskandari, Khoshsima & Safaie-qalati (2021) demonstrate, there are serious differences in the aptness and acceptability of metaphors between cultures. Since previous studies examining the influence of metaphors on reasoning about crime have only been conducted with English, Dutch, and German participants, further research is indispensable to investigate the influence of metaphors on reasoning on a broader scale.

Next, it cannot be overlooked that various news stories, political discourses, and other experiences related to crime that participants experienced shortly before the study influenced their responses. That is, news report about a particular type of crime might have led participants to suggest certain crime-fighting strategies. For instance, if they heard about financially motivated crimes such as burglaries and theft, that are mainly driven by poverty issues, participants might have been prompted to suggest reform-oriented measures such as increasing the number of job opportunities in the city. In contrast, recent exposure to news about violent crimes that occur across all social classes such as domestic abuse and sexual assault, may have primed participants to propose enforcement-oriented suggestions, as reform-oriented measure such as improving the economy and providing health care serve little purpose to prevent such crimes. Similarly, negative

past experiences with police officers may have discouraged some participants from suggesting enforcement-oriented measures. However, such assumptions are only speculative. Further research is needed to clarify what knowledge and beliefs impact participants' decision making.

In addition, Steen, Reijnierse & Burgers (2014) argue 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 where the metaphor is necessary to understand the problem. For instance, regarding complex in mathematics and science, "the more abstract, complex, or unfamiliar the topic, the more likely metaphorical reasoning will be employed" (Bougher, 2012, p. 148). However, because crime is a frequently discussed topic that all participants had presumably encountered prior to the study and because the description of crime was highly simplified, it may be that the issue was not complex enough to require metaphorical reasoning to understand the problem, which could explain why we found no evidence supporting the hypothesis that participants' reasoning was influenced by the metaphor. Nonetheless, if that is the case, it raises the question of how such an effect could be found by Thibodeau & Boroditsky (2011), even though their experiment was based on the same descriptions of crime as the present study.

Furthermore, the main difference between the replication and the follow-up study lies in the presentation of the vignettes. Even though the aim of the follow-up study was to investigate the influence of the speaker's reliability, it cannot be denied that the only manipulation of vignettes was the use of different pictures. Rather, the text in the description is slightly adjusted and the different pictures could activate certain associations that cannot be controlled by the researchers. That is, since the unreliable speaker is depicted by an obviously drunk, aggressive looking person who uses profane language, this may have led participants to associate the unreliable speaker with low social power or status and propose different suggestions than would have been the case if the unreliable speaker were illustrated differently. As becomes apparent in some of participants' responses such as the ones shown in (1) and (2), the portrayal of an unreliable speaker by

a drunk, aggressive looking person caused many participants to associate the unreliable speaker with attributions other than his reduced reliability.

(1)   First off[,] it would be nice to have more even minded and less profane people involved. Also[,] people that seem to be less drunk. […]

(2)   Addison should look into the reasons why the crime is occurring and act based off that, not a drunk man's rant.

(3)   […] This man sounds disgruntled and looks drunk. Maybe more programs for alcohol abuse.

That is, the responses imply that the unreliable speaker is seen not only as less credible, but also less persuasive and less knowledgeable about the issue, even though the descriptions of the crime issue were substantively identical.

In contrast, another participant that was confronted with the *hooligan* condition left the following note (4) at the end of the study.

(4)   I hope that the idea is not that the person's credibility is based on his appearance or use of language, because that would just be plain old judgmental bias.

Assumingly, he was deliberately trying to correct the bias and that, despite the vignette, he was intentionally taking description of the unreliable speaker seriously. Therefore, in this case, the design of the follow-up study failed to achieve its goal of making the hooligan appear less reliable.

Similarly, while the picture of a newscaster was chosen to portray a reliable speaker, this also elicited unexpected associations. That is, not all participants considered a newscaster to be the most reliable source for information of crime. For example, consider the response of one participant in (5).

(5)   Get a new reporter! Too many opinion words were used in the report to scare people[.]

Therefore, those assumptions may have influenced the results as well. However, as becomes apparent in the data, the cases shown in (4) and (5) appear to be the only two cases in which the speaker was not associated with the desired level of reliability. As is evident in the results of the rating task, the other participants associated the speakers with different degrees

of reliability as planned. That is, the mean values of the perceived reliabilities r differed noticeably ($r = 4.60$ in the *newscaster* condition and $r = 3.62$ in the *hooligan* condition).

5.3 Unexpected results

Although the focus of this study lies on the four hypotheses, the results contained some unexpected findings, which are addressed below.

First, 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 in Figure 5 for the follow-up study.
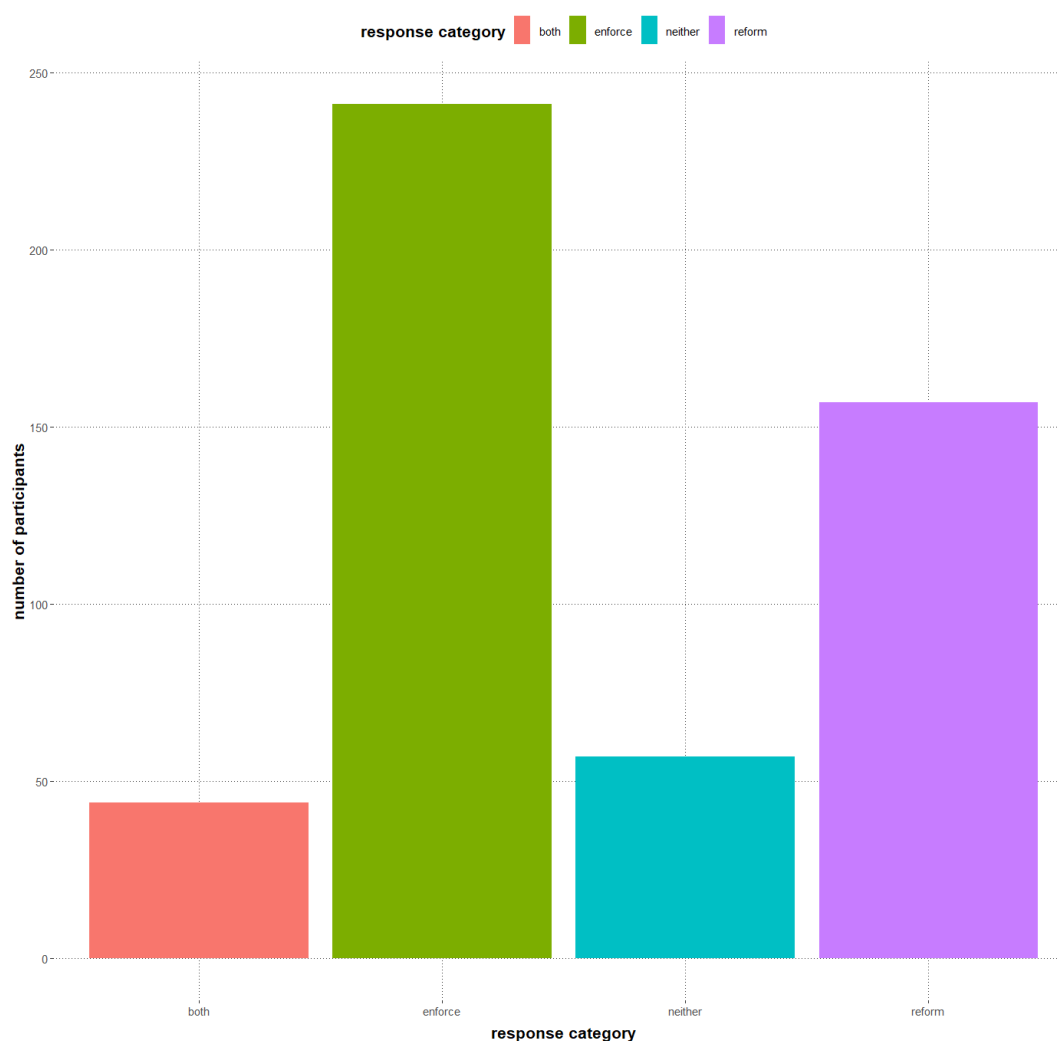

Figure 5: Participants' show overall preference for enforcement-oriented suggestions in the follow-up study.

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 design of the study might therefore have influenced the data, not allowing the results to reach the required threshold.

Second, the distinction between *reform* and *enforce* is unclear in some cases, as the distinction proposed in the original study only distinguishes between measures to address the root causes of crime and measures to punish offenders and does not consider measures that do not hinder offenders from committing crime, e.g., security measures such as increased lighting in public areas and improved security measures in private houses. This raises the question of whether crime deterrence counts as a preventative strategy or an enforcement strategy. It is unclear whether the results would be the same if those responses with the measures in question were categorized differently.

Similarly, according to the original study, investigating root cause should be considered a reform-oriented proposal (cf. Thibodeau & Boroditsky, 2011). However, in their responses, many participants did not suggest examining the root cause of the problem, but rather the nature of crimes committed. It can only be conjectured that exploring the type of the crimes implies that in a next step, the cause of these crimes should also be investigated, but since this is only a conjecture, these suggestions cannot be clearly categorized either.

Additionally, many participants suggested increasing the number of police officers as a strategy to reduce crime. According to Thibodeau & Boroditsky (2011), proposals that focus on police forces should be classified as enforcement-oriented. However, it becomes apparent from the participants' responses that police patrols are seen more as a preventative measure. That is, an increased police presence on the streets of Addison would prevent crime by deterring criminals. For instance, consider responses of two participants shown in (2).

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

(7) 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.

To be more abstract, it can be noted that there seem to be two forms of prevention. First, prevention, in the sense of reform-oriented proposals, includes measures to prevent crime at its root. Second, prevention, in the sense of increased police patrols mentioned by some participants, includes measures that do not curb crime at its root, but deter it out of fear or in anticipation of potential problems or difficulties. Further research could seek to examine a metaphoric framework that leaves less room for interpretation or that explore all different metaphor-consistent and metaphor-inconsistent responses categories.

5.4 Limitations of the study

First, in the present study, recruitment is conducted via the Internet, which is associated with certain biases. For instance, only people who have access to the Internet, who have the required time to fill in the survey, and who are interested in participation are recruited. As a result, certain groups of people may be underrepresented, while others may be overrepresented in the sample. However, conducting the study online perhaps reduced the risk of participants changing their responses because they are being observed by the researchers. This is less likely to be the case in the present study since participants' responses were collected anonymously online.

Second, participants recruitment was based on the original study's procedure. Therefore, the study only includes participants who self-identified as native English speakers based in the United States. However, as pointed out in chapter 5.2, the effectiveness of metaphors differs between cultures. The next steps, therefore, are to replicate the study but including culturally diverse groups of people, exploring their understanding of metaphors in more detail, and attempting to determine the extent to which participants' cultural backgrounds influence their behavior and decision making.

As a last point, our models are based solely on participants' responses and thus, do not account for the effects of prior knowledge and beliefs or sociodemographic information such as age, gender, etc. While we believe

these limitations did not impact the primary outcome of the study, future work may seek to include additional controls.

5.5 Summary

Ultimately, several factors must be considered. On the one hand, the conceptualization of crime could be influenced by a variety of factors such as participants' past experiences, knowledge and beliefs, recent exposure and priming, etc. On the other hand, the influence of metaphor could have been affected by the complexity of the topic, cultural backgrounds of the participants, and assumptions about the speaker. Another difficulty lies in the design of the study, which may have contributed to participants' more assertive suggestions and possibly questionable distinction between the two response categories. Furthermore, the study is subject to limitations caused, for example, by who was recruited and which variables are statistically analyzed. However, as the present study is a replication of Thibodeau & Boroditsky (2011), the design and data analysis had to be similar.

6. Conclusions

Not only the way in which a problem is described, but also external factors can influence how we reason about said problem. In two studies, we investigated the effect of metaphors on reasoning and the contextual factors that influence it. The first study aimed to explore the role of metaphors and political affiliations, while the second study analyzed how the reliability of the speaker impacts 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 of previous studies. Moreover, the results of the second study indicate that (underlying) assumptions of the participants do not seem to heavily influence the effect of metaphors on reasoning. That is, the metaphor framing effect and the influence of participants' assumptions about the speaker might impact their decision but their efficacy appears to

be minimal due to confounding factors and other external factors. Further research is required to explore how metaphors work in more depth.

Gives insights into and Important step towards

- Better understanding of the metaphor framing effect
- Overview of possible(?) influencing factors of the effect
- Insight in the impact of the context

Further research required

- Further statistical analyses that compare the given results to other regression models and casual analyses(?)
- Further research with participants of non-western cultural origin
    - Further research of the influence of speaker's reliability of the metaphor framing effect but with a different metaphor
        - Next steps

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

Eskandari, Z., Khoshsima, H., & Safaie-qalati, M. (2021). Cultural and Cognitive Dimensions of Metaphor Aptness. *Iranian Journal of Applied Language Studies 13*(1), 39-50. https://doi.org/10.22111/IJALS.2021.6398

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

Gibbs, R. W. (2008). *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole, J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Act,* 54-58. Academic Press, New York.

Knowles, M., & Moon, R. (2005). *Introducing Metaphor*. New York: Routledge. https://doi.org/10.4324/9780203642368

Neal, A. A. (2015). *The role of euphemisms in language comprehension: The taboo topic of rape.* Research Thesis. Ohio State University.

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. 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

Teixeira, M. J. O. R. (2022). *Influencer Marketing: Social media influencer's credibility on purchase intention for different levels of product involvement.* Master Thesis. Católica Porto Business School.

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. (2011). Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE*, *6*(2), e52961. https://doi.org/10.1371/journal.pone.0016782

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. (2015). 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., Iyiewaure, P. O., & Boroditsky, L. (2015). 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., & Flusberg, S. J. (In press). *Metaphor and Elaboration in Context*. John Benjamins Publishing. https://osf.io/c93nf/download

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

8. Appendix

8.1 Pilot study: Findings

Figure 6: Plotting reliability against vignette length reveals hardly any influence of vignette length on perceived reliability in pilot study of replication study.
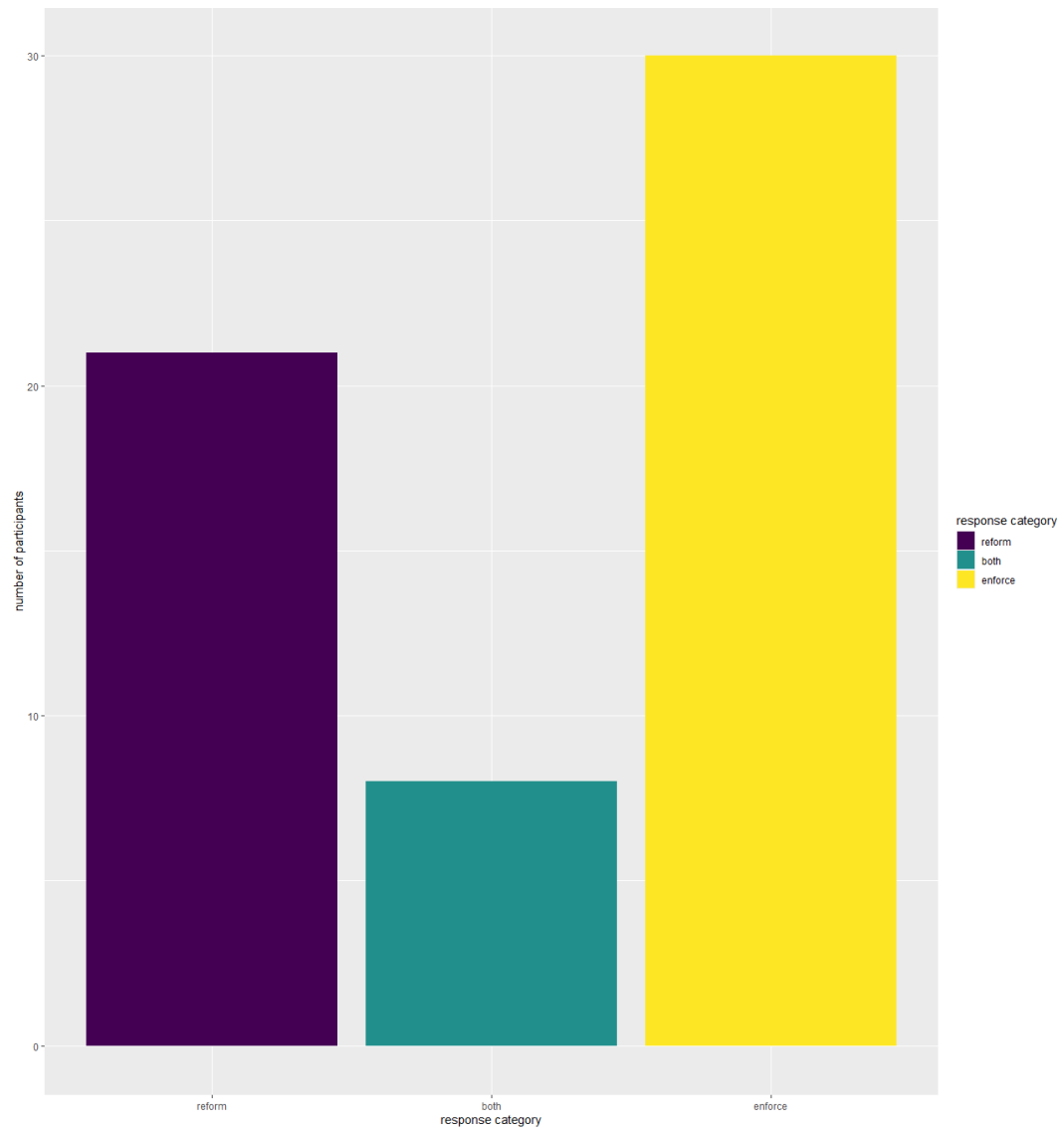
Figure 7: Participants in pilot study of replication study favor enforcement-oriented measures overall.
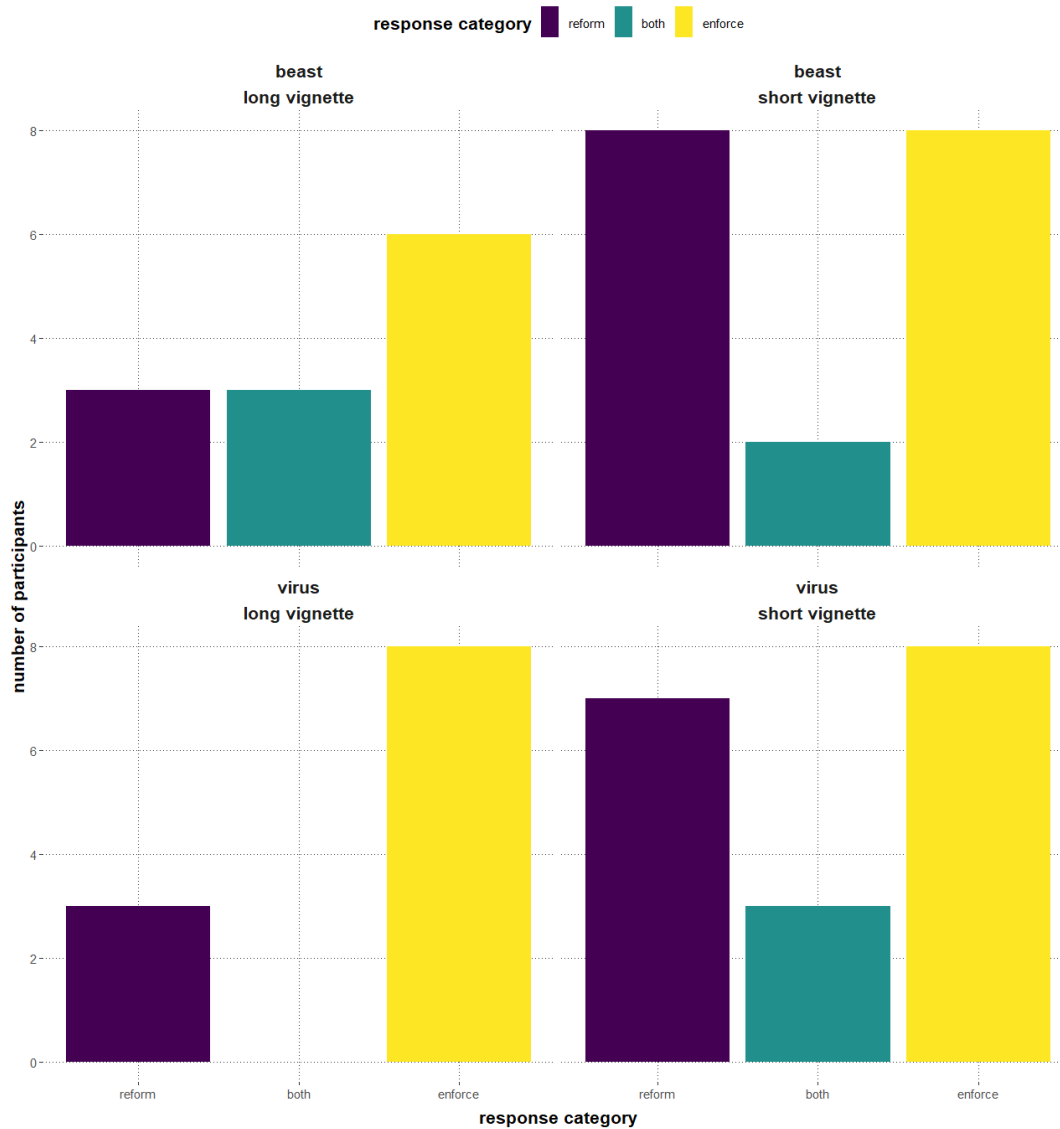
Figure 8: Participants' categorized responses show little impact of metaphorical frame and vignette length in pilot study of replication study.

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.
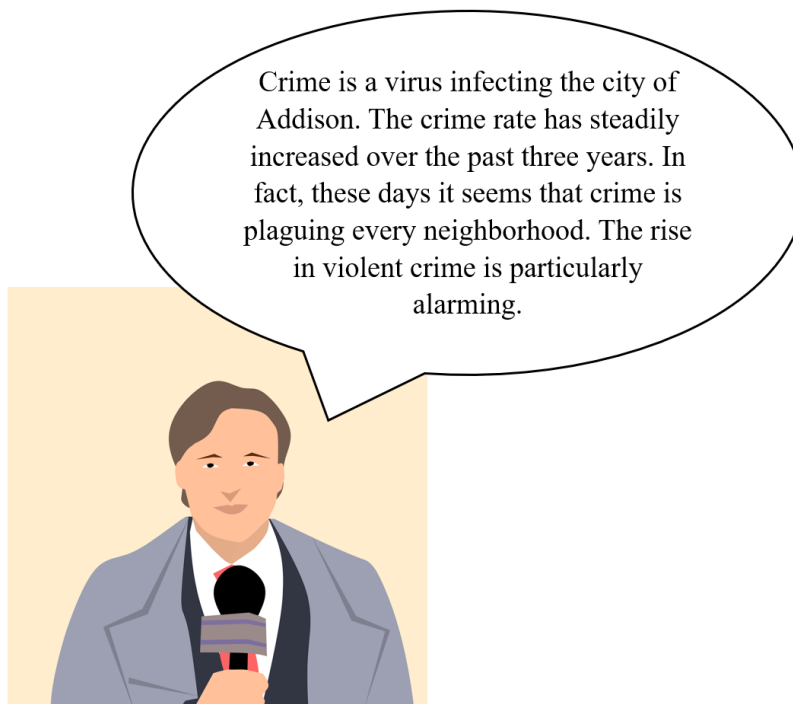
Beast, long vignette:

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.
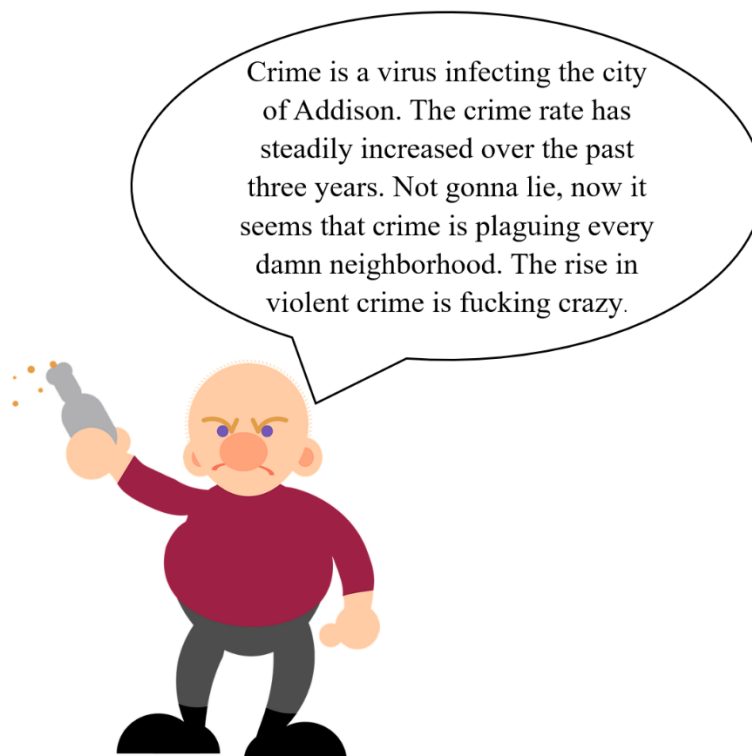
Beast, reliable speaker:

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.
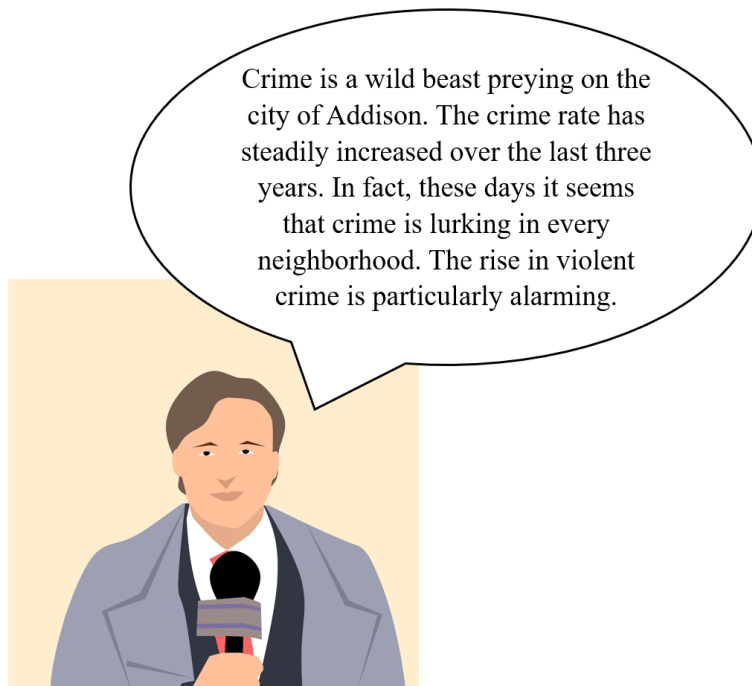
Beast, unreliable speaker:

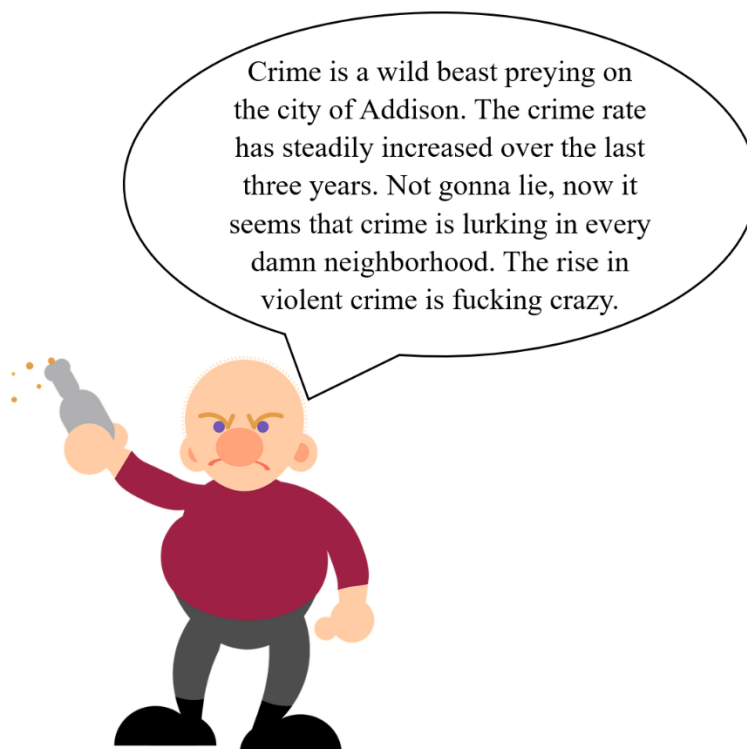Crime is a wild beast preying on the city of Addison. The crime rate has steadily increased over the last three years. Not gonna lie, now it seems that crime is lurking in every damn neighborhood. The rise in violent crime is fucking crazy.

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