CHAPTER 8

Variation in methods for studying political metaphor

Comparing experiments and discourse analysis

Paul H. Thibodeau, James Fleming and Maya Lannen

This chapter explores methodological variation in the study of political metaphor, focusing on a comparison of two approaches: the Critical Discourse Approach (CDA) and experiments (the Response Elicitation Approach; REA). What kinds of political metaphors have been investigated on the two approaches and what insights have these studies revealed? What are the strengths and limitations of each approach? As cognitive psychologists, we have more experience with experiments, and our discussion is grounded in an exposition of the logic and mechanics of experimental design. But we advocate for methodological pluralism because understanding political metaphor is a multifaceted, interdisciplinary endeavor. Some research questions are better addressed through discourse analysis; others are better addressed with experiments; scholars should use the method that is best suited to addressing their research question.

Keywords: metaphor, discourse analysis, research methods, psychology

1. Introduction

Variation is at the heart of political metaphor. This chapter is about variation in two common methods for studying political metaphor. Most of the chapters in this book use a Critical Discourse Approach (CDA; Fairclough 2013; Musolff 2016). With CDA, patterns of real world discourse are identified, quantified, and interpreted to address research questions about political metaphor. One alternative to CDA uses experiments (e.g., Hauser & Schwarz 2015; Landau, Sullivan & Greenberg 2009; Thibodeau & Boroditsky 2013). In a typical experiment on political metaphor, naïve participants read and respond to carefully designed stimuli in a controlled setting. Participants' behavioral responses are analyzed to address psychological questions about how people process metaphors and use them to think. This method has been called the Response Elicitation Approach (REA; Boeynaems, Burgers, Konjin & Steen 2017).

https://doi.org/10.1075/dapsac.85.08thi

In our own work, we primarily use experimental (REA) methods, which means that most of the analyses presented in this book are novel terrain for us. In this chapter we reflect on the relationship between experimental and discourse-based approaches to investigating political metaphor. We start by reviewing a recent study of the similarities and differences of CDA and REA methods (Boeynaems et al., 2017). Boeynaems et al. (2017) compare the types of metaphors that have been investigated and the findings that have been revealed by the two methodological approaches. One of our goals is to step back to consider how the logic and mechanics of the two approaches give rise to the similarities and differences identified by Boeynaems et al. (2017). Because of our experience with experimental methods, and because experiments feature less prominently in the preceding chapters, our discussion is anchored by an exposition of the logic and mechanics of experiments. In other words, we use the experimental approach as a "source domain" for understanding what is for us the comparably novel "target domain" of the discourse-based method.[1] Then we review some of the concerns and limitations of each approach, noting some of the ways that experiments can inform discourse analysis, and some of the ways that discourse analysis can inform experiments.

2. An empirical comparison of discourse-based and experimental approaches to studying political metaphor

One way to compare CDA and REA methods for studying political metaphor is through systematic observation of past experiments and discourse analyses. For example, what are the similarities and differences in the types of metaphors that have been investigated with the two methods? What are the similarities and differences in the findings that have been revealed by experiments and discourse analyses?

A recent study carefully contrasted the two methodological approaches in this way (Boeynaems et al., 2017). A premise of the study is that, "Despite their differences, both aim to answer similar questions, namely how metaphorical framing of political issues affects citizens and society at large" (p. 120). As the authors note, if the two approaches have the same fundamental goal, "It is, therefore, important to empirically examine to what extent these different perspectives report converging or diverging results" (p. 120).

Boeynaems et al. (2017) catalogue 109 previous studies of political metaphor (45 CDA and 57 REA studies). They identify similarities and differences in (a) the types of metaphors investigated and (b) the findings revealed by the two approaches (see Table 1).

1. Again, it is important for us to acknowledge that we have significantly less experience and expertise with the critical discourse approach. Any mistakes or misunderstandings in how we characterize discourse analysis are our own.

Table 1. Similarities and differences in the types of metaphors investigated and the findings revealed by REA and CDA approaches to studying political metaphor, as identified by Boeynaems et al. (2017)

	Similarities	Differences
Metaphor types	Both focus on WAR metaphors.	CDA studies focus on more negatively valenced issues than REA studies.
	Both focus on anthropomorphizing metaphors.	CDA studies focus on more realistic uses of metaphor than REA studies.
	Both focus on dehumanizing metaphors.	CDA studies focus on more intense uses of metaphor than REA studies.
Findings	No noteworthy similarities in findings.	CDA studies always reveal an effect of the metaphor; REA studies do not.
		CDA studies always reveal an effect in line with the frame; REA studies do not.
		CDA studies reveal bigger effects than REA studies.

First, they compare the types of metaphors that have been investigated with the two methods. They find that WAR metaphors, anthropomorphizing metaphors, and dehumanizing metaphors have been studied extensively on both approaches (see Boeynaems et al., 2017 for examples). In other words, there is similarity in the metaphorical source domains that CDA and REA scholars have investigated with respect to political metaphor.

They also find differences in the types of metaphors that have been investigated: "differences in frame valence, frame fictionality, and frame extremity" (p. 124). Frame valence refers to the emotional connotation of the metaphor. CDA studies have focused primarily on negatively valenced metaphors (e.g., negative: "The *war* on terror"; Bartolucci 2012), whereas REA studies have focused on both negatively and positively valenced metaphors (e.g., positive: "The human immune system is an *army*"; Jansen, Van Nistelrooij, Olislagers, Van Sambeek & De Stadler 2010).

Frame fictionality refers to the extent to which the framed issues, "have occurred in real life, and are known by many people" (p. 126). CDA studies focus on real world uses of metaphor in discourse for real world issues. REA studies often test the persuasive influence of metaphors for partially fictional issues – like a crime problem in a fictional city (e.g., Thibodeau & Boroditsky 2011).

Frame extremity refers to how the metaphors are used. Metaphors can be used to express hyperbolic positions, as in "political and social adversaries are parasites" (Musolff 2014), and to frame especially intense events like the 9/11 terror attacks in the United States (Bartolucci 2012). But metaphors can also be used to describe less intense events like employment and tree planting campaigns (e.g., Ahn, Kim & Aggarwal 2014; Robins & Mayer 2000). CDA studies have primarily focused on analyzing the effects of metaphorically framing intense issues. REA studies, on

the other hand, have analyzed the effects of metaphorically framing high-intensity issues, as well as less intense issues.

Second, Boeynaems et al. (2017) compare the results yielded by CDA and REA studies. They find relatively few similarities, concluding that, "differences in reported effects were predominant" (Boeynaems et al. 2017: 127). Whereas discourse analyses always seem to reveal an extreme effect of a metaphor in the predicted direction, the results of experiments are more equivocal: "We noticed differences in effect presence, with all CDA frames being effective, contrary to a part of the REA frames. Furthermore, we saw that CDA effects were always in line with the frame, contrary to REA effects. Finally, we noticed that CDA studies typically reported on effects that were more intense and impactful than REA studies" (Boeynaems et al. 2017: 129). They speculate that some of these differences in findings may be tied to differences in the types of metaphors being investigated (i.e., the emphasis on more extreme, negatively valenced, and realistic metaphors in CDA studies).

Finally, Boeynaems et al. (2017) conclude their study with recommendations for future research. They encourage experimentalists to use "more 'real' and natural language stimuli" (p. 130). They encourage discourse analysts to engage more deeply at the level of the individual, "since language is produced and interpreted in the brain, and the construction of knowledge about social issues thus takes place in the minds of [individual] people" (p. 131). They also recommend studying the time course of metaphor framing effects (e.g., how lasting is the influence of a metaphor frame?) and the boundary conditions of metaphor framing effects (e.g., what factors moderate the persuasive influence of a metaphor?).

We see tremendous value in comparing the two methodological approaches in this way. It is useful to know how scholars have used the two approaches to study political metaphor. We also see wisdom in the recommendations for future research: for experimentalists to investigate more realistic uses of metaphor, for discourse analysts to focus on the individual, and for both approaches to address novel questions about, for example, the time course and boundary conditions of metaphor framing effects.

On the other hand, we also think it is useful to take a step back and reflect on why the two approaches are similar and different. Contrary to a premise of Boeynaems et al. (2017)'s study, that "Despite their differences, both [CDA and REA studies] aim to answer similar questions, namely how metaphorical framing of political issues affects citizens and society at large" (p. 120), we think that the questions being investigated with the two approaches are sometimes fairly different. In turn, there are often good reasons to, for example, use fictional stimuli in experiments and to focus at the societal level in discourse analysis. To put some of these reasons in perspective, we discuss the underlying logic and mechanics of experiments and discourse analysis in the following sections before revisiting Boeynaems et al. (2017)'s findings and recommendations.

3. The logic and mechanics of an experiment

3.1 The components of an experiment

Experiments are comprised of a hypothesis that can be tested, an independent variable that can be manipulated, a dependent variable that can be measured, and a task that yields data (see Figure 1). An experimental hypothesis is a prediction about a causal relationship between variables. For example, one recent experiment tested the hypothesis that WAR metaphors are useful for highlighting the urgency of a sociopolitical issue like climate change (Flusberg, Matlock & Thibodeau 2017).[2] The researchers tested this hypothesis with a task in which participants read a description of climate change and then rated the urgency of the issue. There were three levels of the independent variable: the description of climate change was infused with WAR metaphors, RACE metaphors, or no metaphors (see Table 2). The researchers manipulated the independent variable by randomly assigning participants to one of the three conditions. The dependent variable, urgency, was measured by asking people to respond to a question with a rating on a 5-point scale: "How urgent is it for the US to implement energy reduction programs right away?" The ratings were analyzed to see whether the experiment confirmed or disconfirmed the hypothesis. In this case, participants in the WAR metaphor condition rated climate change as more urgent than participants in the other conditions, thus confirming the hypothesis.

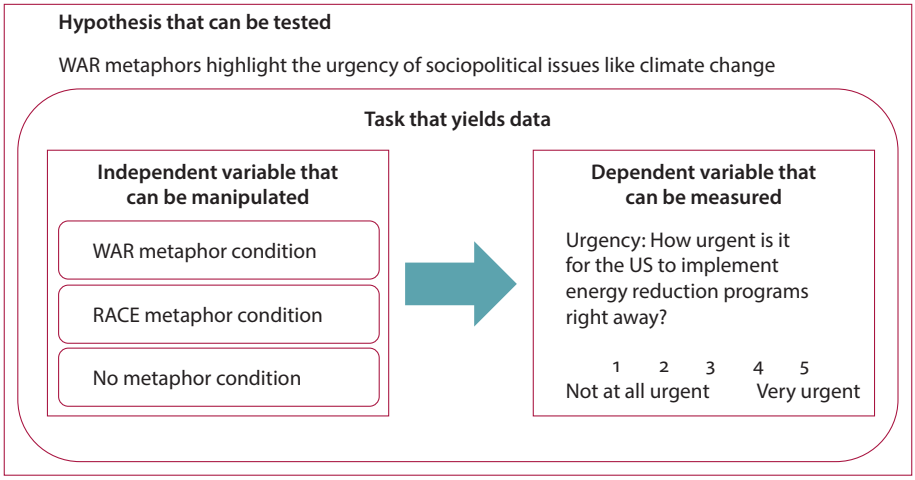

Figure 1. Components of an experiment with examples from a recent experiment on the persuasive influence of metaphor in climate change communications (Flusberg, Matlock & Thibodeau 2017)

2. Note that the full details of this experiment are simplified for didactic purposes.

Table 2. Excerpts of stimuli for three framing conditions of a recent experiment on the persuasive influence of metaphor (Flusberg, Matlock & Thibodeau 2017). Italics added to highlight differences between conditions

War condition	"When will Americans start to *combat* excessive energy use and *kill* the problems related to air pollution and the destruction of natural resources?"
Race condition	"When will Americans *go after* excessive energy use and *surge ahead* on problems related to air pollution and the destruction of natural resources?"
No metaphor condition	"When will Americans start to *address* excessive energy use and *resolve* the problems related to air pollution and the destruction of natural resources?"

3.2 Experimental choices

The validity of an experimental finding hinges on choices made by experimenters and the logic of experimental designs, and choices involve trade-offs. In this experiment, choices were made about how to operationalize the independent variable, how to measure the dependent variable, and how to design the task. For example, the researchers chose the three levels of the independent variable: WAR metaphor condition, RACE metaphor condition, and no metaphor condition. This choice affects the types of inferences that can be drawn from the study. It would be a mistake to infer that WAR metaphors are the *most* effective way to elicit a sense of urgency about climate change because the experiment only contrasted three conditions.

The researchers also chose how to operationalize each of the three levels of the independent variable. The stimuli were inspired by real-world discourse. Newspapers and magazines often use WAR metaphors in headlines and articles about climate change, as in: "We don't need a 'war' on climate change, we need a revolution" (Godoy & Jaffe 2016). They also use RACE metaphors, as in: "In race against climate change, innovations to this ingredient could determine the future of brewing" (Nurlin 2016). However, the text that was used in the study was written by the researchers. And it was written so that the paragraphs expressed parallel information (i.e. they were written so that the only difference between the paragraphs was in the metaphors). In the real world, there may be a tendency to use WAR metaphors to frame one aspect of climate change and RACE metaphors to frame another aspect of climate change. Thus, the way the paragraphs were written and how the metaphors were used represent choices that affect the generalizability of the results. It would be a mistake to infer that WAR metaphors *always* lead people to feel a sense of urgency about the issue they are used to describe because WAR metaphors can be used in many different ways.

Finally, the researchers chose how to design the task and how to measure urgency. It is, of course, possible to measure perceptions of urgency in different ways – with a different question than the one asked in the experiment or by measuring

a different behavior altogether. The researchers could have asked participants to make a charitable donation to a conservation effort. Maybe participants would be more likely to donate after reading about the WAR on climate change compared to the RACE against climate change. That would be a different way to measure urgency – and it would license different inferences (about urgency defined in terms of charitable giving rather than a rating).

3.3 Experimental obligations

Experiments also make certain demands of the researcher. First, they require that the researcher describe and justify the choices that they make. In the study above, the researchers explain why they compared a WAR metaphor condition to a RACE metaphor condition and a no metaphor condition: to see whether the urgency elicited by the WAR metaphors was the result of WAR metaphors *per se* or the result of using metaphoric language in general. If more urgency had been elicited by the WAR and RACE metaphors compared to the no metaphor condition, it might suggest that metaphors, in general, tend to elicit more urgency (or engagement) than non-metaphoric language. However, the experiment found that WAR metaphors elicited more urgency than RACE metaphors, suggesting that there is something particular about WAR metaphors that elicit a sense of urgency.

Second, experiments require that hypotheses are falsifiable – that the task could yield data that disconfirms the prediction. The hypothesis that WAR metaphors elicit more urgency than RACE metaphors in a description of climate change is falsifiable because the experiment could have yielded the opposite result: ratings of urgency could have been higher in the RACE metaphor condition.

Third, experiments require that the hypothesis be specified before data is collected. A hypothesis is a prediction about the future, given a specific set of conditions. Ideally, once a falsifiable hypothesis is specified and an experiment is designed, the researcher's hands are tied. The data are prioritized over the goals of the experimenter. When a prediction is confirmed with an experiment, we gain confidence about the causal relationship between variables. We gain confidence that the prediction will come true again in the future, given a similar set of conditions.

In this way, experimentally confirmed predictions are different from post hoc explanations. A post hoc explanation is an attempt to describe a relationship between variables from data, omitting the step of making and testing a prediction. In turn, the predictive value of post hoc explanations is less clear. For example, maybe there is a correlation between the use of WAR metaphors to describe climate change and feelings of urgency about climate change (e.g., maybe people who are passionate about the issue tend to use WAR metaphors to describe it). If this were the case, it would be hard to know whether WAR metaphors are a cause or a

symptom of urgency. That is, maybe WAR metaphors cause people feel that climate change is an urgent issue; or maybe the feeling that climate change is an urgent issue causes people to use WAR metaphors (or both). Experiments involve manipulating one variable and measuring another in order to test a prediction about the causal relationship between the variables. Specifying the prediction in advance makes the data more informative.

3.4 Summary

In sum, experiments involve choices and trade-offs. On the one hand, they give control to the experimenter to create an artificial environment in order to test causal relationships between variables. On the other hand, they take control away from experimenters – also in the service of testing causal relationships between variables. Experimenters control the hypothesis they test, the way that the variables are defined and manipulated, and the task used to measure them. But they do not have control over the outcome of the data. Experiments are designed so that objective measurements, taken under specified conditions, can be made by anyone.

These choices and trade-offs affect the inferences that can be made from experiments. They enable researchers to draw conclusions about causal relationships between variables under specific conditions. For example, WAR metaphors (operationalized in a particular way) cause people to feel more urgency about climate change than RACE metaphors (in at least some circumstances). As evidenced by the parentheticals in the previous sentence, experiments are also limiting. Based on the results of the experiment described, we can make a prediction about how people will respond to a description of climate change infused with WAR metaphors (with a sense of urgency) – but it can be hard to know the extent to which the results of a controlled experiment generalize to the real world. That is, it can be hard to know about the external validity of an experiment. We return to this issue later in the chapter.

4. The logic and mechanics of discourse analysis

4.1 Similarities to experiments

Discourse analysis is similar to and different from the experimental method. The components of a discourse analysis are similar to those of an experiment: a hypothesis/question, an independent variable, a dependent variable, and a task/situation that yields data.

Like experimenters, discourse analysts make decisions about what questions to investigate and how to measure variables. Table 3 shows the questions investigated in the chapters of this book. For example, Ahrens (Chapter 1) investigates whether the role of a political actor (independent variable) affects their rhetorical style (dependent variable) – by quantifying the use of WAR metaphors in speeches given by Hillary Clinton from 1992 to 2008, during which time Hillary Clinton was the first lady of the United States, a senator, and a presidential candidate. Similarly, Vogiatzis (Chapter 5) investigates a set of questions related to how positive and negative metaphors are used as framing devices to describe difficult sociopolitical situations – by focusing on five major speeches given by the Prime Minister of Greece in 2010–2011. The decisions in these discourse analyses have to be explained and justified, just as they would in an experiment, so that other researchers can evaluate the findings and, if they choose, replicate or extend the results.

Table 3. Research questions explored in this book

Chapter	Research question
(1) Ahrens	Does the role of a political actor affect their rhetorical style?
(2) Heyvaert	Does the political stability of a country affect the prevalence of deliberate metaphors?
(3) Borcic & Culo	How do politicians use metaphors of personification to construct their public image as a leader?
(4) Fenton-Smith	How do politicians use certain metaphor scenarios (leadership as direction, conflict, foundation, hard work) to describe leadership takeovers?
(5) Vogiatzis	How do politicians use positive and negative metaphors to frame difficult sociopolitical situations?
(6) Ströbel	How do candidates for political office use sensorimotor-based concepts to persuade and influence the general public?
(7) Kovář	What do patterns of conceptual metaphor usage reveal about a political party's position about an issue like EU finality?

As with experiments, the choices made in the process of conducting discourse analysis affect the inferences that can be drawn from the study. Ahrens finds systematic variation in Hillary Clinton's use of WAR metaphors that corresponds to the public roles Clinton served in her political career. This suggests that a person's political role affects their rhetorical style. However, additional research is needed to test whether the specific patterns for Clinton generalize to other political actors. Vogiatzis finds that the Prime Minister of Greece extensively used positively valenced metaphors to frame the austere economic policies that the country was forced to adopt, suggesting an important role for metaphor in establishing the emotional tone of a

message. Again, however, additional research is needed to test whether the specific patterns found in this analysis generalize to other political leaders talking about other challenging sociopolitical situations.

Finally, the process of investigating research questions in discourse analysis shares some similarities with hypothesis testing in experiments. Hypotheses (or exploratory research questions) guide decisions about data collection on both methodological approaches. For example, Ahrens develops and analyzes a corpus of speeches given by Hillary Clinton to test a hypothesis about the relationship between *political role* and *rhetorical style*. Similarly, Vogiatzis collects and analyzes a particular set of speeches given by the Greek Prime Minister in 2010–2011 to explore how metaphors are used to frame unpopular economic policies.

4.2 Differences from experiments

There are also important differences between the experimental approach and the discourse-based approach. First, there are subtle differences in the components of the two approaches and the choices that researchers make. The "task" that yields data for discourse analysis is different from an experimental task. In an experiment, the researcher creates a controlled environment to isolate the relationship between the variables of interest (e.g., read one of three descriptions of climate change and rate urgency in an online survey). The independent variable is manipulated, and the dependent variable is measured.

With discourse analysis, the researcher finds creative ways to isolate the relationships between variables by focusing on specific types of real-world situations: situations in which the independent variable changes naturally and the dependent variable can be measured (e.g., Charteris-Black 2011; Fairclough 2013; Musolff 2016). For example, Ahrens focuses on speeches given by Hillary Clinton at different points in her career to operationalize *political role*; and analyzes the content of those speeches to measure *rhetorical style*. Similarly, Vogiatzis focuses on speeches given by the Greek Prime Minister about a specific topic – the financial crisis – and analyzes the content of those speeches to measure the use of positive and negative metaphors. In both of these examples, patterns between variables are revealed by choosing the situation from which the data are analyzed, rather than by manipulating one variable and measuring another.

Second, there are differences in how hypotheses are investigated on the two approaches. Experimental tasks are tailored to the hypotheses they are designed to test. Experimenters can tweak and rerun experiments to collect new data. If one experiment yields a surprising result, then a new hypothesis can be specified, a new task can be designed, and a follow-up experiment can be carried out. With

discourse analysis, hypothesis testing is more constrained. We can replicate the analysis of Hillary Clinton's speeches from 1992 to 2008, but not the conditions that yield the data (i.e., we cannot go back in time and have Clinton give new speeches). As a result, there are also differences in the types of hypotheses and research questions that can be investigated on the two approaches. Experiments are better suited to testing causal relationships between variables in a controlled setting. Discourse analysis is better suited to identifying and characterizing relationships between variables in the real world.

Third, the methods often engage at different levels of analysis. Experiments typically focus at the level of the individual (e.g., how do individuals process metaphors and use them to think?). Discourse analyses often focus at a more collective or societal level (e.g., how does a political leader's rhetorical style affect their popularity and ability to enact a political agenda in a country?).

Fourth, there tend to be differences in how data are analyzed on the two approaches, and, in turn, the types of knowledge they produce. Experiments are typically designed to yield quantitative data that are analyzed with statistical tests. In the experiment described earlier, the researchers found that reading a description of climate change infused with WAR metaphors elicited a greater sense of urgency ($M = 4.34$, $SD = 0.99$) than a description of climate change infused with RACE metaphors ($M = 4.25$, $SD = 0.95$), $t\ (2182) = 1.98$, $p = .048$. In this case, the statistical analysis tells us is that there is a 4.8% chance that the result is a "false positive" (i.e. that WAR and RACE metaphors actually elicit similar levels of urgency, even though the study found that WAR metaphors elicit more). Note that 4.8% is unlikely, but possible. In this way, the inference that WAR metaphors, compared to RACE metaphors, *cause* people to feel a greater sense of urgency about climate change is probabilistic. It is licensed by the logic of the experimental design – that the only difference between the WAR and RACE conditions was in the metaphors – and the statistical analysis.

Discourse analysis can involve quantitative measurement and statistical analysis – as Ahrens's and Vogiatzis's chapters show. However, they often involve analyzing qualitative patterns of language use. For example, Fenton-Smith (Chapter 4) discusses the role of four metaphor scenarios (LEADERSHIP as DIRECTION, as CONFLICT, as FOUNDATION, and as HARD WORK) in party leadership takeover speeches. Discussion of these metaphor scenarios and inferences about the roles that they play in party leadership takeover speeches are supported with examples from a corpus. Such descriptions of qualitative patterns of language use yield a different kind of knowledge than is produced by experiments. This type of knowledge is not inherently better or worse than the knowledge produced by experiments, but it is different.

Differences in the types of knowledge produced by the approaches, in some cases, reflect different epistemological commitments of the two approaches. Epistemology is the philosophical study of knowledge: What is it? How do we attain it? What human capacities make knowledge possible? (Amundson 1985). Experimentalists are largely positive empiricists, viewing knowledge as something that can be gained through systematic observation – by asking the right questions and taking valid measurements. This is why experimentalists create controlled environments to measure variables, why they emphasize hypothesis testing, and why they describe knowledge as probabilistic. A different epistemological commitment underlies at least some discourse analysis: a social constructivist view, in which knowledge is constructed and maintained through social practice. On this view, the value of knowledge comes from its applicability, usefulness, and clarity (Potter 1996). This is one reason why discourse analyses emphasize the real world, rather than controlled settings, as a source of data, and why discourse analyses often involve greater discussion of qualitative patterns of language use, rather than quantitative measurement and statistical tests.

4.3 Summary

There is a similarity in the components, logic, and mechanics of the experiments and discourse analysis. In both, researchers find ways to isolate and make inferences about the relationships between variables through systematic observation.

On the other hand, the two approaches often ask different types of research questions, find different ways to isolate and make inferences about the relationships between variables, and, in turn, produce different types of knowledge, which are grounded in different epistemological commitments and focused at a different level of analysis. Experimenters typically ask targeted questions about causal relationships between variables that can be manipulated and measured in a controlled environment with individuals. Discourse analysts often ask questions about relationships between variables in society at large that can be observed in a naturalistic setting.

5. A return to empirical similarities and differences

Thinking through the underlying logic and mechanics of experiments and discourse analysis can help contextualize the empirical similarities and differences of the two approaches, as identified by Boeynaems et al. (2017). For example, both approaches have investigated WAR metaphors, humanizing metaphors, and dehumanizing metaphors because these metaphors are pervasive in political discourse

(Flusberg, Matlock & Thibodeau 2018; Gibbs 1994; Lakoff & Johnson 1980; Musolff 2016). That is, both approaches are informed by real world patterns of language use.

Differences in the fictionality of the metaphors that have been investigated with the two approaches can be attributed to differences in the goals and logic of the two approaches. The goal of discourse analysis is to understand real world patterns of language use (i.e. to study realistic uses of metaphor). Experiments, on the other hand, aim to isolate relationships between variables in a controlled setting. They are fictional by design. For example, in the experiment described earlier, the researchers tested whether WAR metaphors elicit more urgency than RACE metaphors with carefully constructed paragraphs that were designed to express parallel information (i.e. to differ only in metaphoric language). The WAR metaphor condition could have been more hyperbolic as well. That would make it more realistic. But then it would be hard to know whether a difference in urgency between the conditions resulted from using WAR metaphors *per se* or from the use of hyperbole (or a combination of the two factors). In this sense, the real world applications of the experiment are secondary. The primary goal is to address a psychological question about how people process and interpret different metaphors for climate change. Using fictional stimuli in experiments reduces the potential influence of confounding variables.

The ability to create fictional stimuli allows experimentalists to explore a wide range of metaphors (both more and less extreme; both positively and negatively valenced). Discourse analyses have focused on more extreme and negatively valenced metaphors because these are some of the most interesting and influential real world instances of political metaphor.

The goals and methods of the two approaches also provide a context for thinking about why they yield different results. Experiments are designed to test specific hypotheses and to produce probabilistic knowledge. The observation that experiments sometimes find surprising results is heartwarming to experimenters – it's a feature, not a bug. Sometimes experiments yield surprising results for the same reason that a fair coin will, on rare occasion, land heads up ten times in a row (that is just how probability works). Other times experiments yield surprising results because the hypothesis was wrong. In this way, surprising results help researchers refine their theories – they help us gain new knowledge. If every experiment simply confirmed the prediction of the experimenter, how would we ever learn anything new? Indeed, an experimentalist might be cautious about the method of discourse analysis because the results seem to be too good to be true (Casasanto 2009; Keysar & Bly 1995; McGlone 2011).

Discourse analyses, on the other hand, are often designed to show and explain why a real-world event unfolded the way that it did: how, for example, patterns of

metaphor use change as a political actor takes on different roles; how a political leader uses metaphor to set an emotional tone to explain an unpopular policy or a dramatic transition of power. This is why the results of discourse analysis tend to find stronger, more consistent effects than experiments. The situations and patterns of metaphors investigated with discourse analysis are chosen deliberately – because they reveal something interesting about the relationship between metaphor, politics, and society.

Of note, we have encountered a number discourse analyses that highlight failed uses of political metaphor. For example, Musolff (2016) shows how political metaphors can have dramatic, unintended effects. In 1991, British Prime Minister J. Major proclaimed that Britain would, "work 'at the very heart of Europe'" (p. 40; originally from *The Guardian*, 12 March 1991). The metaphor implies that it is good to be at the center of a decision-making body, and that it is possible for Britain to be at the center of the European Union. However, this phrase was turned against its original intended meaning – because many Britons thought it would be impossible or impractical for Briton to serve a central role in the EU. Indeed, many Britons thought it could be harmful – it could cause a "coronary" (p. 43, Musolff 2016; originally from the *Economist*, 26 September 1992).

In sum, it is important to acknowledge that the research questions, methods, and types of knowledge produced by experiments and discourse analysis are similar in some ways and different in others. Simply comparing the metaphors investigated on and results produced by the two approaches can obscure fundamental similarities and differences in the underlying goals, logic, and mechanics of the methods.

6. Limitations of experiments and discourse analysis: A crossroads, with avenues for future research

Concerns have been raised about experimental and discourse-based approaches to studying political metaphor. In some cases, the limitations of the two approaches complement one another. For example, a major critique of experiments is that they are overly reductionist (e.g., Boeynaems et al. 2017; Charteris-Black 2011). Experiments are designed to test causal relationships between variables by manipulating one and measuring another. And in order to manipulate and measure variables, experimenters have to use controlled (artificial) tasks. This limits the types of questions that can be addressed with experiments, and it raises questions about the generalizability (external validity) of the results produced by experiments. For example, it is not clear to what extent WAR metaphors in actual news articles about climate change affect perceptions of the urgency of climate change in the real

world. We can make a prediction based on the results of an experiment, but the knowledge that we gain from an experiment is necessarily limited.

On the other hand, a major critique of discourse analysis is that it is not reductionist enough (e.g., Casasanto 2009; McGlone 2011). Discourse analysis relies on real world patterns of language use. It is a correlational method. The analyst finds natural variation in one variable and measures another to see if the two are related. This raises questions about the nature of the relationship between the variables: X could cause Y; Y could cause X; or Z could cause X and Y. For example, Ahrens' finds that Hillary Clinton's *political role* is correlated with aspects of her *rhetorical style*. But it is unclear whether changes in Clinton's *political role* cause differences in her *rhetorical style* based on the data we have – because there are a lot of confounding variables. In addition to changing political roles between 1992 and 2008, Clinton got older, the US engaged in a literal war on terror, and there were dramatic changes in the social, political, and cultural context of the United States. It is hard to draw inferences about causal relationships between variables with correlational methods.

We agree that these two critiques are valid; however, we do not find them particularly concerning. Partly this is because what is framed as a limitation of the experimental approach is a strength of discourse analysis (and vice versa). More importantly though, we believe that research questions should guide the methodological choices that researchers make. Basic psychological questions about how people interpret metaphors, produce metaphors, and use metaphors to think are well suited to test with experimental methods. Psychological experiments inform psychological theories; and psychological theories have a proven track record of producing useful knowledge (e.g., Kahneman 2011). Nuanced questions about real world political rhetoric are well suited to discourse analysis. Such analyses inform theories at the intersection of politics, culture, language, and society writ large; and these theories also have a proven track record of producing useful knowledge (e.g., Charteris-Black 2011; Gibbs 1994; Lakoff & Johnson 1980; Musolff 2016).

There are, of course, opportunities to use experiments to test insights gained through discourse analysis. Indeed, a number of experiment ideas emerge from the analyses presented in this book. For example, do interpretations of political metaphors vary as a function of the speaker's political role? An experiment could manipulate the supposed source of a speech (e.g., "Hillary Clinton said this when she was first lady vs. a senator vs. a presidential candidate") and measure participants' interpretation of the speech. Similarly, an experiment could test how people respond to austere policy speeches that are infused with positive versus negative metaphors; and an experiment could test how people respond to different political leadership takeover speeches that are infused with different metaphor scenarios.

There are also, of course, opportunities to use discourse analysis to test insights gained through experiments. Some are simple. For example, do environmental activists gravitate to specific types of metaphor scenarios to highlight the urgency of climate change? Some are more nuanced. For example, recent experimental work in psychology has given substantial attention to a range of cognitive, social, and affective factors that make some metaphors more persuasive than others (Thibodeau, Hendricks & Boroditsky 2017). One could imagine comparing the types of metaphors used by various political actors to see whether some leaders use metaphors more effectively than others; and whether the effective use of metaphor is correlated with real world outcomes – like the leaders' popularity or their ability to enact a political agenda.

Nevertheless, we think it is important to recognize that there are inherent limitations to both approaches. A common refrain is for experimenters to use more realistic stimuli in their experiments in order to "enhance the ecological validity of their research" (p. 131; Boeynaems et al., 2017). As we have noted, we see wisdom in this suggestion, especially if the goal of an experiment is to generate knowledge that is immediately useful for an applied real world purpose. For example, a researcher might want to compare two climate change messaging strategies in order to see which is more likely to make people donate money to a conservation effort. In this type of hypothetical experiment, controlling for differences between the messages would be less important, since the goal is to find the message that yields the most donations. On the other hand, such an experiment would be less informative to someone interested in psychological theory, since the two messages would likely be different in a number of ways. And if the two messages are different in a number of ways, it is hard to infer why one message caused people to donate more than the other.

In short, there is no perfect experiment, nor is there a perfect discourse analysis. Both involve choices and trade-offs. Both should be designed in a way that best addresses the research questions they seek to answer.

7. Conclusion

The collective message of this book is that there is variation in political metaphor: variation in how metaphors are used, what they mean, and how they function. What we have emphasized in this final chapter is that there is also variation in how metaphors are investigated. We have focused on experiments and discourse analysis. We have tried to think through the logic and mechanics of the two approaches in a way that will help experimentalists understand discourse analysis (and vice versa).

In our view, both methods have distinct strengths and limitations. Mapping the logic of the approaches onto one another helps to highlight these strengths and limitations – for researchers and consumers of research. Experiments are designed to test hypotheses about the causal relationships between variables in a controlled setting. But not all variables can be manipulated in a controlled setting, so not all research questions can be addressed with experiments. And it can be hard to generalize from an experiment to the real world. Discourse analysis, on the other hand, is designed to investigate patterns of relationships between variables in the real world. But variables cannot be manipulated and measured in the real world like they can in an experiment, so it can be hard to know exactly how the variables are related to one another.

Embracing the strengths and limitations of the two approaches can help us develop even more useful theories about the nature of political metaphor. They can help us learn even more about the nature of politics, culture, society, and language by studying political discourse.

References

Ahn, H. K., Kim, H. J. & Aggarwal, P. (2014). Helping fellow beings anthropomorphized social causes and the role of anticipatory guilt. *Psychological Science*, 25, 224–229. https://doi.org/10.1177/0956797613496823

Amundson, R. (1985). Psychology and epistemology: The place versus response controversy. *Cognition*, 20, 127–153. https://doi.org/10.1016/0010-0277(85)90050-2

Bartolucci, V. (2012). Terrorism rhetoric under the Bush administration: Discourses and effects. *Journal of Language and Politics*, 11, 562–582. https://doi.org/10.1075/jlp.11.4.05bar

Boeynaems, A., Burgers, C., Konijn, E. A., & Steen, G. J. (2017). The effects of metaphorical framing on political persuasion: A systematic literature review. *Metaphor and Symbol*, 32, 118–134. https://doi.org/10.1080/10926488.2017.1297623

Casasanto, D. (2009). When is a linguistic metaphor a conceptual metaphor? *New Directions in Cognitive Linguistics*, 24, 127–145. https://doi.org/10.1075/hcp.24.11cas

Charteris-Black, J. (2011). *Politicians and Rhetoric: The Persuasive Power of Metaphor*. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230319899

Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. New York: Routledge. https://doi.org/10.4324/9781315834368

Flusberg, S. J., Matlock, T., & Thibodeau, P. H. (2017). Metaphors for the war (or race) against climate change. *Environmental Communication*, 11, 769–783. https://doi.org/10.1080/17524032.2017.1289111

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

Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

Godoy, E. S., & Jaffe, A. (2016, October 31). We don't need a 'war' on climate change, we need a revolution. *The New York Times*. Retrieved November 3, 2016, from http://www.nytimes.com/2016/10/31/opinion/we-dont-need-a-war-on-climate-change-we-need-a-revolution.html?_r=0

Hauser, D. J., & Schwarz, N. (2015). The war on prevention: Bellicose cancer metaphors hurt (some) prevention intentions. *Personality and Social Psychology Bulletin*, 41, 66–77. https://doi.org/10.1177/0146167214557006

Jansen, C., Van Nistelrooij, M., Olislagers, K., Van Sambeek, M., & De Stadler, L. (2010). A fire station in your body: Metaphors in educational texts on HIV/AIDS. *Southern African Linguistics and Applied Language Studies*, 28, 133–139. https://doi.org/10.2989/16073614.2010.519102

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, 34, 89–109. https://doi.org/10.1006/jmla.1995.1005

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Landau, M. J., Sullivan, D., & Greenberg, J. (2009). Evidence that self-relevant motives and metaphoric framing interact to influence political and social attitudes. *Psychological Science*, 20, 1421–1427. https://doi.org/10.1111/j.1467-9280.2009.02462.x

McGlone, M. S. (2011). Hyperbole, homunculi, and hindsight bias: An alternative evaluation of conceptual metaphor theory. *Discourse Processes*, 48, 563–574. https://doi.org/10.1080/0163853X.2011.606104

Musolff, A. (2014). Metaphorical parasites and "parasitic" metaphors: Semantic exchanges between political and scientific vocabularies. *Journal of Language and Politics*, 13, 218–233. https://doi.org/10.1075/jlp.13.2.02mus

Musolff, A. (2016). *Political Metaphor Analysis: Discourse and Scenarios*. New York: Bloomsbury.

Nurin, T. (2016, August 31). In race against climate change, innovations to this ingredient could determine the future of brewing. *Forbes*. Retrieved November 3, 2016, from http://www.forbes.com/sites/taranurin/2016/08/31/in-race-against-climate-change-innovations-to-this-ingredient-could-determine-the-future-of-brewing/#733b53417810

Potter, J. (1996). *Representing Reality: Discourse, Rhetoric and Social Construction*. London: Sage. https://doi.org/10.4135/9781446222119

Robins, S., & Mayer, R. E. (2000). The metaphor framing effect: Metaphorical reasoning about text-based dilemmas. *Discourse Processes*, 30, 57–86. https://doi.org/10.1207/S15326950dp3001_03

Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PloS ONE*, 6, e16782. 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.

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Sciences*, 21, 852–863. https://doi.org/10.1016/j.tics.2017.07.001