Spring 2022

# The Role of Force Dynamic Schemas in the Comprehension of Causal Language

Dawson Petersen

# THE ROLE OF FORCE DYNAMIC SCHEMAS IN THE COMPREHENSION OF CAUSAL LANGUAGE

By

Dawson Petersen

Bachelor of Science
North Greenville University, 2019

---

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

Linguistics

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

Amit Almor, Director of Thesis

Anne Bezuidenhout, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

ABSTRACT

The current study is an attempt to empirically test the predictions of Talmy's (1988) force dynamics. Specifically, Talmy argues that causal sentences are understood by reference to basic image schemas, such as Starting and Stopping. While many of the predictions of other cognitive linguistic models, such as Lakoff and Johnson's (1980) conceptual metaphor theory, have been tested empirically by psycholinguists (Fischer & Zwaan, 2008; Gibbs, 2006; Boroditsky & Ramscar, 2002), force dynamics has received very little empirical attention (I am only aware of Wolff & Song, 2003), in spite of its productivity in formal linguistics.

The current study aims to fill this gap by employing a priming paradigm in which participants experience a force dynamic prime followed by self-paced reading of a sentence with a causative verb which referenced a compatible or incompatible force dynamic schema. In order to strengthen the priming manipulation, the primes used were short interactive 2D computer games created using Unity and the Unity Experiment Framework as primes (Brookes et al., 2019). In each game, participants had to either prevent or cause an object to move against the force of an antagonist object. Following the prime, participants read sentences that described events that had the same or the opposite force dynamic schema as the prime as well as neutral sentences that did not describe events with a force dynamic schema. Each sentence was presented region by region, and the response time for each region was recorded. It was predicted that participants would respond more quickly when the target sentence matches the interactive

force dynamic prime. However, the results show no priming effect from force dynamic primes to causal sentences which suggests either that force dynamic schemas are not used in online sentence comprehension or that their influence is weak enough that an even stronger manipulation is required to uncover it.

TABLE OF CONTENTS

Abstract .... iii

List of Tables .... vi

List of Figures .... vii

Chapter 1: Literature Review .... 1

Chapter 2: Experiment One .... 13

Chapter 3: Experiment Two .... 24

Chapter 4: General Discussion .... 29

References .... 37

Appendix A: Experiment One Stimuli Sentences .... 42

Appendix B: Experiment Two Stimuli Sentences .... 44

Appendix C: Additional Analyses .... 47

# LIST OF TABLES

Table A.1 Critical Sentences Used in Experiment One ....................................................42

Table A.2 Nonsense and Filler Sentences Used in Experiment One ................................43

Table B.1 Critical Sentences Used in Experiment Two ....................................................44

Table B.2 Nonsense and Filler Sentences Used in Experiment Two ................................45

LIST OF FIGURES

Figure 2.1 The Syntactic Structure of the Linguistic Stimuli ..........................................15

Figure 2.2 Example Starting Stimuli ....................................................................17

Figure 2.3 Example Stopping Stimuli....................................................................18

Figure 2.4 Critical Region Residual Reading Times by Condition ..................................21

Figure 2.5 Post-Critical Region Residual Reading Times by Condition..........................22

Figure 3.1 Critical Region Residual Reading Times by Condition ..................................26

Figure 3.2 Post-Critical Region Residual Reading Times by Condition..........................27

# CHAPTER 1

# LITERATURE REVIEW

The goal of the current study is to empirically test the predictions of two theoretical approaches within cognitive linguistics, namely Talmy's (1988) force dynamics and Lakoff and Johnson's (1980) conceptual metaphor theory. In Chapter 1, I will review the relevant literature focusing on the theoretical background from cognitive linguistics, as well as the psycholinguistic research on conceptual grounding, in order to provide reading time predictions based on Talmy's (1988) force dynamics. In Chapters 2 and 3, I will describe two experiments which tested the predictions outlined in Chapter 1. Finally, in Chapter 4, I will discuss the results of these experiments and their implications for cognitive linguistic accounts of language processing.

## 1.1 COGNITIVE LINGUISTICS

Cognitive linguistics is a broad approach to semantics and syntax rooted in the assumption that the structure of language is grounded in the structure of cognition (Evans & Green, 2006). Therefore, cognitive linguists are interested in the ways in which conceptual categories can be used to explain syntactic and semantic structure. Cognitive linguistics distinguishes between two sets of conceptual categories (Talmy, 1988). Specifically, it is argued that some conceptual categories, such as number and gender, which are frequently encoded in closed-class lexical categories and syntax, play a basic role in the structuring of language, while other categories, such as color, do not (Talmy, 1988).

### 1.1.1 FORCE DYNAMICS

One of these proposed grammatical conceptual category is force dynamics, which Talmy (1988) argues can be used to explain causative language, especially modals, as well as certain aspects of prepositions and causative verbs. As a conceptual category, force dynamics covers not only prototypical CAUSING, but also concepts such as LETTING, PREVENTING, and HELPING. Talmy (1988) suggests that the structure of force dynamics is as follows.

Firstly, Talmy argues that the conception of force as manifested in language assumes that entities have intrinsic tendencies towards motion or rest, which can be illustrated with examples such as 1) (1988).

1) a. The ball was rolling along the green.

b. The ball kept rolling along the green.

Talmy argues that 1a) is force dynamically neutral. While 1a) depicts movement, it carries no information as to how forces are interacting. In contrast, 1b) makes the notion of force salient. Specifically, for 1b) to be true from the standpoint of intuitive physics, one of two things must be the case. Either the ball has an internal tendency against motion (i.e., inertia) which is being overcome by an external force (such as the wind), or the ball has an internal tendency to roll (for example, due to gravity) which is being resisted by an external force (for example, from stiff grass) (1988).

In addition, as is suggested by examples such as 1b), when two force dynamic entities are opposed, they do not receive equal linguistic (and, Talmy would argue, conceptual) treatment. Under either interpretation of 1b), the ball is privileged over the other force dynamic entity, which in this case is entirely omitted (though it can be included using a prepositional phrase such as “because of the wind” or “despite the thick grass”). It should be noted that this privileging is independent of the tendency or relative strengths of the entities. Although under the first reading the ball is receiving force and in the second it is generating force, it remains in a privileged position. Talmy refers to the privileged entity as the agonist and its partner as the antagonist (1988). Furthermore, it should be noted that this difference is not purely a result of the syntax. For example, if the antagonist is put into the subject position, as in 2), the agonist still cannot be omitted.

2) a. The wind kept the ball rolling.

b. *The wind kept rolling.

Given these assumptions about the structure of the force dynamic concepts, force dynamic situations can be naturally divided along three lines: 1) the tendency of each entity towards rest or motion, 2) the relative strength of the entities, and 3) the duration of the interaction, namely whether it is continuous or punctual. Each unique combination of these factors results in a unique schema. Consider the following examples (Talmy, 1988).

3) The ball kept rolling because of the wind blowing on it.

4) a. The log kept lying on the incline because of the ridge there.

b. The shed kept standing despite the gale wind blowing on it.

c. The ball’s hitting it made the lamp topple from the table.

3) and 4a) differ in the assumptions about the tendency of the agonist and antagonist. 3) exemplifies prototypical causation. It is assumed that without the external force towards motion from the wind, the ball would remain stationary. It contrast, 4a) exemplifies prevention. It is understood that if the ridge were not preventing it, the log would naturally roll down the incline. 3) differs from 4b) in the relative strength of the

agonist and antagonist. Unlike in 3), in 4b) the agonist is stronger than the antagonist, allowing the agonist to resist the external force towards motion. Finally, 3) differs from 4c) in duration. While 3) indicates a continuous steady-state force dynamic situation while 4c) indicates a punctual change-of-state dynamic.

On Talmy's view, these same force dynamic schemas are important for the conceptual structure of both naïve physics and folk psychology (and, by extension, the linguistic structure of both concrete and abstract causatives). Talmy presents the following example (1988).

5) a. She is polite to him.

b. She is civil to him.

Like 1a-b), 5a-b) form a minimal pair in which 5b) represents a force dynamic schema while 5a) does not. While both 5a-b) encode the idea of politeness, 5b) contains additional information, specifically that she is polite to him in spite of some internal resistance to doing so. In force dynamic terms, this can be categorized as a steady-state schema in which an agonist with a tendency towards movement is resisted by a weaker antagonist. In other words, she has a tendency to continue being polite to him which prevails in spite of her competing desire to be rude.

On the linguistic level, there is a strong argument for the existence of force dynamic schemas from the grammatical category of modals. Talmy (1988) presents the examples in 6).

6) a. John cannot leave the house.

b. John would not leave the house.

c. John need not leave the house.

Talmy argues that these modals naturally carve up the conceptual space of force dynamics. In each of these cases, *John* is an agonist being pitted against an unknown antagonist, but the tendency of each and the relationship between them differs when the modal is changed. 6a) invokes a schema in which the agonist has a tendency towards movement, but a stronger antagonist restrains it. 6b) invokes a schema in which the agonist has a tendency towards rest and is not compelled to move despite pressure from a weaker antagonist. Finally, 6c) represents a schema in which a weaker agonist with a tendency towards rest has been released of pressure towards movement by a stronger antagonist. Although not all modals are this neatly described in force dynamic terms (for example, the difference between *may* and *can* has to do with the animacy of the antagonist), these modals provide compelling evidence that force dynamic schemas are grammaticalized linguistic entities. However, as I will show in section 1.2, the psychological reality of force dynamics as a conceptual tool has not yet been demonstrated.

### 1.1.2 CONCEPTUAL METAPHOR THEORY

While force dynamics predicts that the same schemas underlie both concrete and abstract cognition about causation, Talmy (1988) does not provide any specific account as to how this abstraction takes place. One potentially informative explanation of this abstraction comes from conceptual metaphor theory (Lakoff & Johnson, 1980). Like Talmy, Lakoff and Johnson are interested the relationship between linguistic and cognitive structures. Specifically, they argue that the systematic metaphors in conventional language are a result of systematic analogical cognitive structures which they term conceptual metaphors (1980). They give the following examples (italics original):

7) a. Your claims are *indefensible*.

b. He *attacked every weak point* in my argument.

c. His criticisms were *right on target.*

The examples in 7) are all highly conventionalized metaphors (to the point that they may even be mistaken for literal language); however, when examined, metaphors like these display a high degree of systematicity. In this case, each of the metaphors in 7) implicitly frames argument in terms of war. Lakoff and Johnson argue that these examples are all instantiations of the conceptual metaphor ARGUMENT IS WAR (1980). Many more conceptual metaphors have been identified, and they provide a great deal of explanatory power on a purely linguistic level by revealing coherent systematic connections between different instantiations of figurative language. However, Lakoff and Johnson go further to argue that conceptual metaphors are not merely fossils of conventional language, but rather living members of the cognitive system. There are several theoretical arguments in support of this claim.

Firstly, these conceptual metaphors show a consistent one-directional pattern. Consider the following examples.

8) a. feeling down (to mean feeling sad)

b. ?being sad (to mean being at a low elevation)

9) a. to be fuming (to mean being angry)

b. ?to be angry (to mean emitting smoke)

10) a. We'll do that ahead of Thursday (to mean prior to Thursday)

b. ?This runner is prior to that one (to mean ahead of them in a race)

In each of the metaphor pairs in 8-10), a) is a conventional expression and b) is its reversal. There are two interesting facts to note about examples like these. Firstly, while the a. examples are readily interpretable, the b) examples rarely are. Secondly, the a) examples show a general pattern of concrete to abstract. In order to explain these patterns, which persist throughout conventional language, Lakoff and Johnson suggest that conceptual metaphors are a way of grounding abstract cognition in more concrete domains (1980). This proposal would explain not only why these systematic metaphors exist, but also their one-directional nature and the overwhelming tendency towards moving from concrete source domains to abstract target domains.

In addition, the productivity of these conceptual metaphors provides evidence that they are active in cognition. For example, English utilizes the conceptual metaphor IDEAS ARE OBJECTS, as evidenced by the examples in 11).

11) a. She *gave* me the idea.

b. She *has* many good ideas.

c. She *brought* that idea *to the table.*

Lakoff and Johnson note that this metaphor can be extended productively in may novel ways, giving the examples in 12).

12) a. juggling ideas

b. dressing ideas up in fancy clothing

c. lining up ideas nice and neat

While the expressions in 12) are far less conventional than those in 11) (and as such, much more likely to be noticed as metaphorical), they are nevertheless readily interpretable because of their relationship to the conceptual metaphor IDEAS ARE OBJECTS. Together, the directionality of metaphors, along with their productivity, provide evidence, at least from a theoretical perspective, that conceptual metaphors play an active role in language processing and cognition more generally.

If Lakoff and Johnson's proposal that concrete concepts play a role in the structuring of abstract concepts is correct, it could help to explain why causal image schemas based on naïve physics would play a role in linguistic processing of more abstract cognition about causation, such as cognition about psychological and social causation.

## 1.2 GROUNDING

From a psycholinguistic perspective, both force dynamics and conceptual metaphor theory can be seen as instances of conceptual grounding. While conceptual grounding is a well-studied topic and the role of metaphor specifically in grounding has received a lot of attention (for example Thibodeau & Boroditsky, 2011; Gibbs, 2008), the grounding of causal reasoning has been much less studied.

### 1.2.1 SENSORIMOTOR GROUNDING

One domain in which Lakoff and Johnson's (1980) predictions have been tested extensively is embodiment, i.e. the grounding of cognition in the sensorimotor system. In major literature reviews, Fischer and Zwaan (2008) and Gibbs (2006) argue that the sensorimotor system plays an important role in the comprehension of both concrete and, via metaphor, even abstract language.

This claim is based on evidence from both the behavioral and neurological literatures. For example, Creem and Proffitt (2001) had participants grasp various objects while saying the names of objects with different grasping affordances. They demonstrated that this semantic task significantly inhibited grasping performance. Similarly, Glover et al. (2004) showed that the aperture of participants hands early in grasping motions could be affected by hearing the name of an object which requires either a large or a small grip. Together, studies like these show an overlap between semantic and motor processing which can be explained as a grounding of semantic concepts in motor simulations. Glenberg and Kaschak (2002) showed further evidence for this explanation by showing the existence of an action-sentence compatibility effect, meaning that listening to a sentence about a hand or foot action inhibited incompatible responses with the hands or feet respectively. Buccino et al. (2005) further demonstrated, using TMS, that participants showed motor evoked potentials in the hands and feet when listening to hand and foot sentences respectively.

Further evidence for overlap between semantic and motor processing for concrete language can be found in the neurological literature. Aziz-Zadeh et al. (2006) replicated Buccino et al. (2005)'s findings using fMRI. Aziz-Zadeh et al. (2006) found that when participants read literal sentences about hand, foot, and mouth actions, participants showed significant activation in the relevant areas of the motor cortex. To test the extent to which this effect relies on motor familiarity, Calvo-Merino et al. (2006) had ballet dancers watch videos of dance which either matched or mismatched the subject's gender. Because ballet includes many moves which are performed only by men or women but not both, this meant that, while for the matched moves, participants would have a great deal of motor experience, for the mismatched moves, participants would have visual experience but little to no motor experience. Calvo-Merino et al. (2006) found fMRI evidence of motor simulation for the matched but not the mismatched moves. Studies like these help to support the claim of overlap between sensorimotor and semantic systems, at least for concrete actions and concepts.

Further research has tested Lakoff & Johnson's (1980) prediction that these effects should extend to metaphorical language (for a comprehensive review, see Jamrozik et al. 2015). One study by Gibbs, Gould, and Andric (2006) showed that when participants performed or imagined a bodily movement before hearing a related metaphor, they produced more elaborate explanations and were more likely to refer to bodily actions in their explanation. A follow-up study by Wilson and Gibbs (2007) found that the same interventions shortened reading times for related metaphors compared to a control. Similarly, Glenberg and Kaschak (2002) showed that the action-sentence compatibility effect held not only for sentences about object transfer (e.g. "Liz gave you the toy") but also for sentences about information transfer ("Liz told you the story"), even though the information transfer sentences used in this experiment were not explicitly metaphorical. Glenberg and Kaschak (2002) argue that this finding is consistent with Lakoff and Johnson's (1980) analysis according to which words are metaphorically viewed as containers for information which can be passed from a speaker to a listener. A follow-up study by Glenberg et al. (2008) replicated these effects for both object and information transfer sentences, and further demonstrated using TMS that these transfer sentences resulted in more hand motor impulses than control sentences. Overall, there is strong evidence, both behavioral and neurological, that many semantic representations (even those of abstract concepts) are grounded in the sensorimotor system.

### 1.2.2 TEMPORAL GROUNDING

Another example of grounding is the relationship between temporal and spatial concepts. Boroditsky (2000) argues that many temporal concepts are grounded in and structured by spatial concepts. As evidence for this, Boroditsky compares two incompatible spatial metaphors for time—the ego moving (for example "the revolution is *before* us") and time-moving metaphors (for example "the revolution was over *before* breakfast"). She showed that when participants saw a movement diagram that referenced an ego-moving perspective, they would be primed to interpret an ambiguous sentence using the ego-moving metaphor and vice versa. Later work has shown that the same priming effect can occur after experiencing, or even just thinking about, travel (Boroditsky & Ramscar, 2002), and that switching between ego- and time-moving metaphors was associated with processing cost (Gentner, Imai, & Boroditsky, 2002). Overall, these results provide evidence that temporal reasoning is metaphorically grounded in spatial reasoning. However, as I will show in the next section, little has been done to investigate the extent to which abstract reasoning about causation is based on more concrete domains.

### 1.2.3 CAUSAL GROUNDING

Like conceptual metaphor theory, force dynamics has had an impact on theoretical linguistics. However, very little experimental work has been done to test the psychological reality of force dynamic schemas or to test how abstract instantiations of these schemas might be grounded. One study by Wolff and Song (2003) compared force dynamics to an alternative theory, the focal set model of causal cognition, which argue that causal concepts are abstracted from probabalistic observations rather than drawn directly from naïve physics as in force dynamics (see Cheng & Novick, 1992). Specifically, Wolff and Song had participants sort causal verbs into clusters and argued that the patterns of sorting were more consistent with the focal set model than with force dynamics. However, this study only examined conscious judgements about the relationships between verbs, not online processing. In addition, this study did not examine the role of image schemas in the comprehension of causal language. An unpublished study by Madden and Pecher (2010), which was reported in Pecher, Boot, and Van Dantzig (2011), did attempt to investigate the psychological reality of force dynamic schemas. Madden and Pecher had participants watch animations which referenced force dynamic schemas, and then collected reading times for sentences referencing matching or mismatching schemas. However, after several experiments, this study did not produce compelling evidence for the psychological reality of force dynamic schemas (Pecher, personal communication, May 3, 2021). Therefore, there is still an important gap in the literature since, firstly, it is not clear whether, and to what extent, image schemas are used in the processing of concrete and abstract causal sentences, and, secondly, if the same image schemas are used in both cases, it is not clear how this abstraction takes place. The current study attempts to fill this gap by advancing on Madden and Pecher's methodology by using interactive games rather than animations as primes. Since interactive primes will force participants to think and solve problems involving force dynamics, rather than passively observing them, this should result in a stronger manipulation and be more likely to uncover an effect if one exists.

# CHAPTER 2

# EXPERIMENT ONE

## 2.1 DESIGN

The first experiment is designed to test the basic prediction of Talmy's (1988) force dynamic model that causal language comprehension is based on causal image schemas. If causal language comprehension is based on image schemas, then it should be expected that exposure to these schemas should facilitate the processing of compatible causal sentences and inhibit the processing of incompatible sentences. The current study investigates this question using two schemas—a Starting schema (in Talmy's terms a change-of-state schema with a weaker agonist with a tendency to rest) and a Stopping schema (a change-of-state schema with a weaker agonist with a tendency to motion). In each trial, participants first completed a game in which they either had to start or stop an object moving and then read a sentence which was compatible, incompatible, or neutral to the paradigm in the game. As an attention check, nonsense sentences, which participants were required to reject, were included in some trials.

## 2.2 PARTICIPANTS

In order to test this hypothesis, this study recruited 95 healthy right-handed, native-speakers of English. Participants were recruited via Amazon Mechanical Turk. The rate of compensation was adjusted over the course of data collection in order to increase participation and varied from $2 to $3.50. Participants who did not complete the experiment (n = 13) were not compensated, and their data were excluded from analysis, resulting in a sample size of 82. Of these, five participants fell below an accuracy threshold of 80% and were excluded, resulting in a final sample size of 77. Each participant completed anonymous electronic consent form as well as a basic demographic survey prior to completing the experiment. Participants were informed that they were not required to complete the study if they become unwilling to at any point, that they would not be exposed to any unusual risks, and that all data would be collected anonymously. The only data collected was response-time, correctness data, and basic demographic information to ensure that the participants were eligible to participate in the study.

## 2.3 MATERIALS

### 2.3.1 LINGUISTIC STIMULI

Fifty sentences were generated to be used in a self-paced reading paradigm. Of these, twenty referenced a force dynamic schema; ten were force dynamically neutral; and the remaining twenty were nonsense sentences to be used as attention checks. Among the critical trials, half exemplified a Starting schema and half exemplified a Stopping schema.

In order to reduce variability between items, each of the stimuli sentences had the syntax shown in Figure 2.1. Some of the stimuli were adapted from Talmy's (1988) examples, while others were created originally. However, all of the adapted examples were altered to match the required syntax.

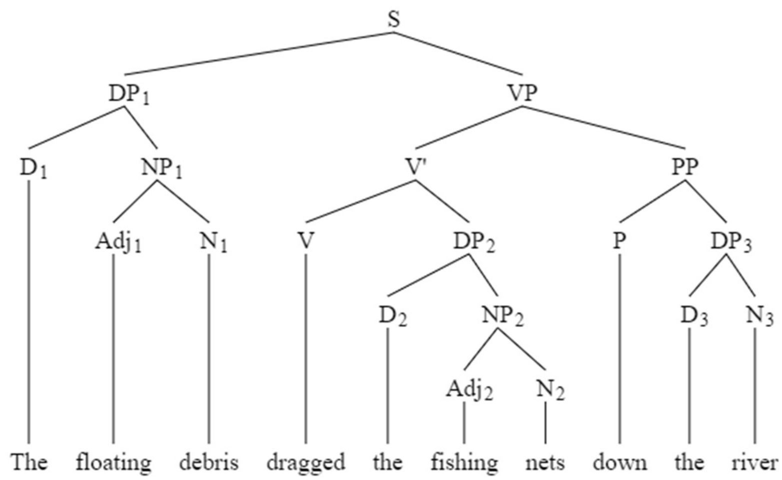


Figure 2.1 The Syntactic Structure of the Linguistic Stimuli

Verbs were selected based on three primary factors: concreteness, force dynamic content, and preferred syntax. As to concreteness, concrete verbs were preferred over abstract verbs to overly complex semantics. As to force dynamic content, verbs were selected to precisely Talmy's Starting and Stopping schemas, rather than causation and prevention more generally. For example, Talmy distinguishes between continuous prevention and punctual prevention (Stopping), so care was taken to ensure that verbs like *kept* which suggest continual prevention were not used since they would be less consistent with the priming stimuli. Finally, each verb was checked against the syntactic frame in Figure 2.1 and excluded if it was incompatible or overly awkward in that frame. Nouns were selected for concreteness and lack of agentiveness again to avoid semantic complications. All of the NPs were definite. The Starting and Stopping sentences were composed as equivalents with the only difference being whether the agonist or the antagonist was stronger (see Appendix A). Each of the sensible sentences depicted physical movement, with the critical sentences involving interaction between entities while the neutral sentences involved the movement of one entity in reference to two other non-moving entities.

In order to ensure constant attention, the nonsense sentences were constructed to stop making sense at different points throughout the sentence—early, middle, and late. The early sentences involved semantic incompatibility between the subject and verb, often due to an agentive verb and a non-agentive subject. The middle sentences involved incompatibility between the verb and the object. Finally, the late sentences involved improbable prepositional phrases.

The linguistic stimuli were pilot tested in a self-paced reading paradigm with four participants. Any sentence which was falsely rejected by any of the participants was reexamined and edited to insure comprehensibility.

### 2.3.2 INTERACTIVE STIMULI

The current study required the development of a new methodology for priming causal image schemas. Specifically, since the goal of these stimuli was to promote causal reasoning in order to activate causal image schemas, a stronger manipulation, compared to previous studies such as Madden and Pecher (2010), was achieved by having participants interact in a causal system rather than passively observing one. Therefore, the priming stimuli consist of 50 short, abstract 2D games created using Unity and the Unity Experiment Framework (Brookes et al., 2019). In each game, participants control one object using the arrow keys and are tasked with using the object they control to manipulate a second object. Like the sentence stimuli, the interactive priming stimuli are designed to represent punctual events of physical causation, namely the causation of motion (Starting) and the causation of rest (Stopping). Each of the stimuli took between 5

- 15 seconds to complete. Figures 2.2 and 2.3 show examples of Starting and Stopping stimuli respectively.

In each of the Starting stimuli, participants were presented with a scene containing a red block (which the player controlled), a blue ball, and a gold flag. Participants were instructed during training that when they saw a gold flag that they should push the ball to the flag. In order to create some variation in the stimuli, these scenes were varied systematically across four irrelevant surface features, namely 1) whether the overall slope to the flag was inclined, 2) whether there was a hill in between the ball and the flag, 3) whether the ball needed to be pushed off the edge of a sharp decline, and 4) whether the direction of movement was from right-to-left or left-to-right.

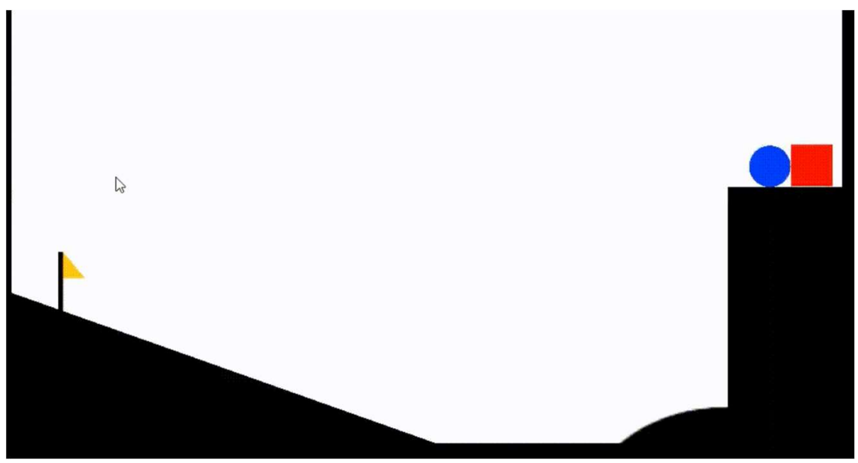

Figure 2.2 Example Starting Stimuli

The Stopping stimuli were presented in the same way, except that a red, rather than gold, flag was presented, and the ball started in such a position that it would naturally roll to the flag unless the participant prevented it from doing so. Participants were instructed during training that when there was a red flag, then they should make the ball came to a complete stop before it reached the flag. If the ball reached the flag, the trial was reset, and the trial ended when the participant kept the ball at a complete stop for

1000 milliseconds. Like the Starting stimuli, these scenes were systematically varied across three surface features, namely 1) whether the ball dropped from a higher level before rolling to the flag, 2) whether the ball started on a suspended platform, and 3) the direction of movement.

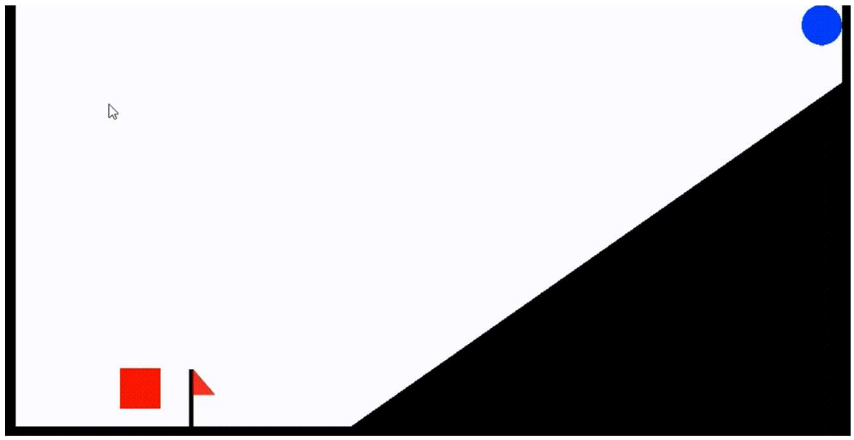

Figure 2.3 Example Stopping Stimuli

2.4 PROCEDURE

Each of the interactive stimuli was paired with one of the sentences such that the critical sentences had an equal number of compatible and incompatible primes, and the pairs of stimuli were presented in a random order. In order to avoid confounds, two versions of the experiment were created such that the primes for the Starting and Stopping sentences are reversed. In other words, the sentences with incompatible primes in the first version of the experiment had compatible primes in the second version and vice versa. The sentences were presented using a moving-window region-by-region self-paced reading paradigm. For each sentence, the precritical region is the subject NP; the critical region consists of the VP up to the object but excluding the PP; finally, the postcritical region includes the concluding PP. The reading time for each region was collected. Based on Talmy (1988), I predicted that there would be an effect of prime-sentence compatibility in the critical and/or postcritical regions.

Participants first completed a brief informed consent and background survey. Then each participant was presented with one of the two versions of the experiment at random. After completing a short tutorial which explained the self-paced reading and interactive stimuli, participants completed ten practice trials, which used only neutral and nonsense sentences. The practice trials were presented in the same ratio of sensible to nonsensible sentences as in the actual experiment (3:2). During the practice trials, but not afterwards, the participants received feedback as to whether or not they correctly identified the sentence's sensibility. Then the participants completed the remaining fifty trials of the experiment in a random order. The entire procedure took approximately 20 minutes. All recruitment and experimental procedures were approved by the University of South Carolina IRB.

## 2.5 RESULTS

The response time data were analyzed using two linear mixed-model regressions with the reading time for the critical (from the verb until the direct object) and post-critical (the ending prepositional phrase) regions as dependent variables respectively. Prior to the main analysis, we analyzed the raw reading times in a regression with region length in characters and word frequency as predictors and then used the residuals from this analysis as the dependent measure in all subsequent analyses.[1] Following residualization, outliers were removed from the data using an adjusted boxplot rule,

[1] In addition to analysis by residuals, the data was analyzed using reading rate (reading time divided by region character length) and raw reading times as dependent variables. See Appendix C.

which accounted for the inherently skewed nature of response time data (Hubert & Vandervieren, 2008). 14.34% of the data from critical trials was excluded in this way. Analysis was performed using lmerTest 3.1.3 (Kuznetsova, Brockhoff & Christensen, 2017) in R 4.0.2 (R Core Team, 2020).

The primary purpose of the analysis was to test the hypothesis that there would be an interaction between prime type and sentence type such that sentences would be facilitated by matching primes, while accounting for the effects of region length and word frequency. In each of the models, the interaction between sentence and prime type was modelled as fixed factors, with participant and item number modelled as random factors. There were two levels of prime type (Starting and Stopping) and three levels of sentence type (Starting, Stopping, and Neutral). Analysis was performed both including the Neutral sentences as a control condition and excluding them to compare the Starting and Stopping conditions more directly. In each model, the maximum random effects structure was used unless the model failed to converge, in which case the model was simplified (Barr et al, 2013).

### 2.5.1 CRITICAL REGION ANALYSIS

Firstly, region reading times were modelled for each participant using region character length as both a fixed and random factor (Trueswell et al., 1994). Then, a new model was fit to the residuals from that model with sentence and prime type and their interaction as fixed factors and participants and items as random factors. This model was fit both with and without the Neutral sentence condition. When the Neutral sentence condition was included, there were no significant interactions, only main effects of sentence type such that the Neutral [$\beta = -256.9$, $t(29) = -2.476$, $p = .019$] and Stopping sentences [β = -260.6, t(35) = -2.828, p = .008] were read more quickly than the Starting sentences. With the Neutral sentence condition excluded, there was a significant effect of sentence type such that the Stopping sentences were read more quickly than the Starting sentences [β = -260.3, t(25) = -3.490, p = .002]. In both models, the critical interaction of prime and sentence type was not significant (p > .05). The critical region residual reading time data is illustrated in Figure 2.4.

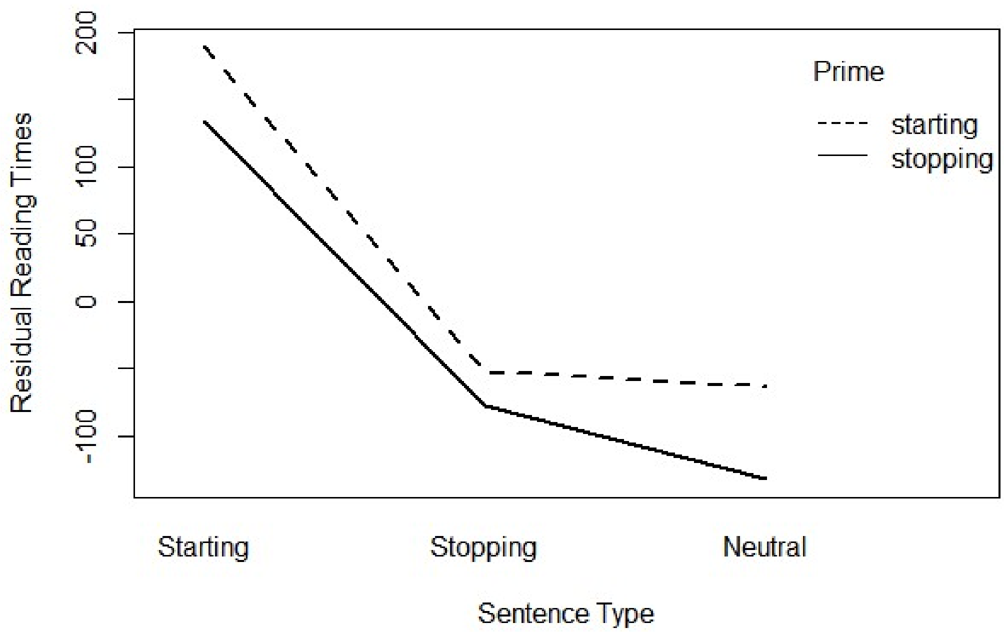


Figure 2.4 Critical Region Residual Reading Times by Condition

## 2.5.2 POST-CRITICAL REGION ANALYSIS

The same analysis was performed for the post-critical region (the ending prepositional phrase). Post-critical region reading times were modelled for each participant using region character length as both a fixed and random factor (Trueswell et al., 1994). Then, a new model was fit to the residuals from that model using the interaction of sentence and prime type as fixed factors, with participants and items as random factors. With the Neutral sentence condition included, there were no significant interactions, only a main effect of sentence type such that Starting sentences were read more slowly than the Neutral sentences [$\beta = -337.6$, $t(24.7) = -2.087$, $p = .047$] and Stopping sentences [$\beta = -348.7$, $t(27.3) = -2.576$, $p = .016$]. With the Neutral sentence condition excluded, there was a significant effect of sentence type, such that the Starting sentences were read more slowly than the Stopping sentences [$\beta = 348.9$, $t(19.24) = 2.260$, $p = .035$]. There was no interaction between sentence and prime type ($p < .05$). The post-critical region residual reading time data is illustrated in Figure 2.5.

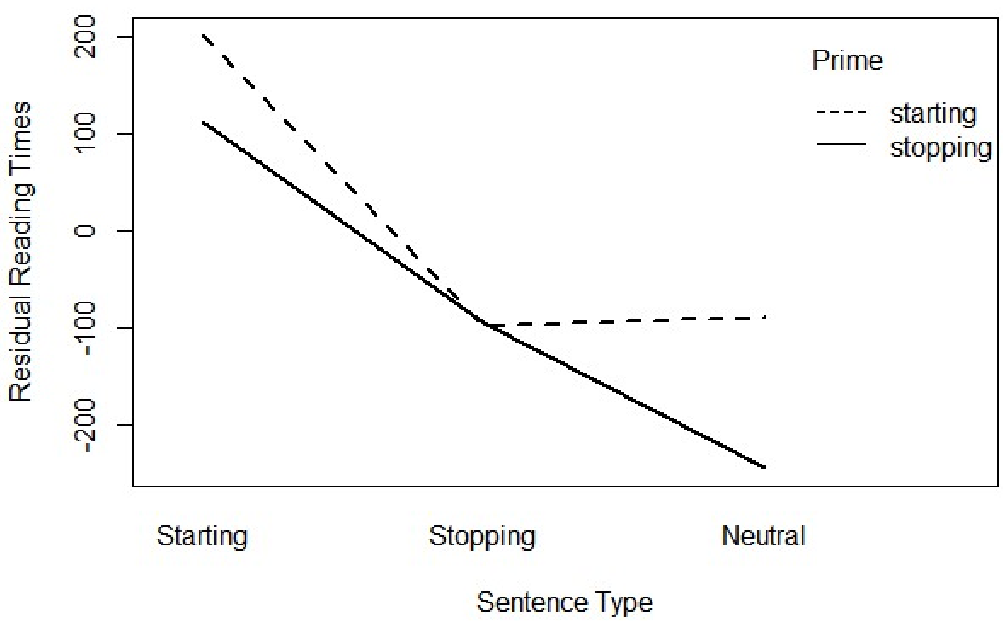


Figure 2.5 Post-Critical Region Residual Reading Times by Condition

2.5.3 POWER ANALYSIS

In order to test whether the lack of an interaction between prime and sentence type could be attributed to a lack of power, a post-hoc simulation-based power analysis was conducted for both the critical and post-critical regions using the simr package 1.0.5

in R 4.0.2 (Green & MacLeod, 2016). In each simulation, 200 samples were generated based on the variance and betas from the linear model. However, the beta of the interaction between task and prime was adjusted to test for power assuming different effect sizes in the population. Assuming a relatively small effect size with a standardized $\beta = .15$, the simulations showed 91% power for the critical region and 91.5% power for the post-critical region. If the true effect size were even smaller at $\beta = .1$, the simulations showed only 66% power for the critical region and 61.5% power for the post-critical region.

## 2.6 DISCUSSION

The goal of experiment one was to test whether image-schematic primes would affect the comprehension of concrete causal sentences, as predicted by Talmy (1988). I predicted that, if image-schematic primes are implicated in the comprehension of causal sentences, that there should be an interaction between sentence and prime type on reading times, such that when sentence type matched prime type that processing would be facilitated either in the critical or postcritical region. The results of experiment one do not support this hypothesis, as no interaction was found in the data. The only significant effects found in the data was that the Starting sentences were consistently read more slowly than the Stopping and Neutral sentences. This difference cannot be explained by word frequency or region length since both were considered in the regression analysis, or by syntactic effects as the syntax was held constant across conditions. Further explanations for this effect are considered in the general discussion. Experiment two will test whether a prime-sentence compatibility effect can be found for abstract-metaphorical psychosocial causation sentences.

# CHAPTER 3

# EXPERIMENT TWO

## 3.1 DESIGN AND PROCEDURE

Following experiment one, a second experiment was designed to test whether a prime-sentence compatibility effect could be found for psychosocial causation sentences with metaphoric verbs. The design and procedure of experiment two were the same as experiment one, except that the sentence stimuli used concrete verbs metaphorically to discuss psychosocial causation.

## 3.2 PARTICIPANTS

94 new participants were recruited and compensated using the same means as experiment one. Of these, twelve did not complete the entire experiment and were excluded, resulting in a sample size of 82. None of these participants fell below an 80% accuracy threshold. All participants in experiment two were compensated at a rate of $3.50.

## 3.3 MATERIALS

The same interactive stimuli from experiment one were used in experiment two. However, new linguistic stimuli were constructed (see Appendix B). The new sentences evoked the same force dynamic schemas and all utilized concrete verbs, reusing the same verbs as experiment one whenever possible. However, the subject in the new sentences depicted abstract entities rather than physical ones, the objects depicted individual people, and the prepositional phrase was replaced with a prepositional gerund phrase as shown in 13).

13) a. The brilliant writing dragged the reluctant schoolboy into reading the novel.

b. The impenetrable writing barred the eager schoolboy from reading the novel.

The Neutral and nonsense stimuli were adapted to also include agents and follow the same syntactic structure as the critical sentences.

## 3.4 RESULTS

The data from experiment two were also analyzed using two linear mixed-model regressions with the residual reading times for the critical (from the verb until the direct object) and post-critical (the ending prepositional phrase) regions as dependent variables respectively.[2] As in experiment one, outliers were removed using an adjusted boxplot rule (Hubert & Vandervieren, 2008). 13.24% of the data from critical trials was removed this way. Analysis was again performed using lmerTest 3.1.3 (Kuznetsova, Brockhoff & Christensen, 2017) in R 4.0.2 (R Core Team, 2020). Analysis was both performed including the Neutral sentences as a control condition and excluding them to compare the Starting and Stopping conditions more directly. In each model, the maximum random effects structure was used unless the model failed to converge, in which case the model was simplified (Barr et al, 2013).

[2] As in experiment one, the data was additionally analyzed using reading rate and raw reading times as dependent variables. See Appendix C.

### 3.4.1 CRITICAL REGION ANALYSIS

As in experiment one, the data from the critical region was modelled using mixed-factor regression. Each model was run both with and without the Neutral sentence condition as a reference.

As in experiment one, region reading times for the critical region were modelled for each participant using region character length as both a fixed and random factor (Trueswell et al., 1994), and then a new model was fit to the residuals from that model using the interaction of sentence and prime type as fixed factors, with participants and items as random factors. This model did not show any significant effects, with or without the Neutral sentence condition ($p > .05$). The residual data is illustrated in Figure 3.1.

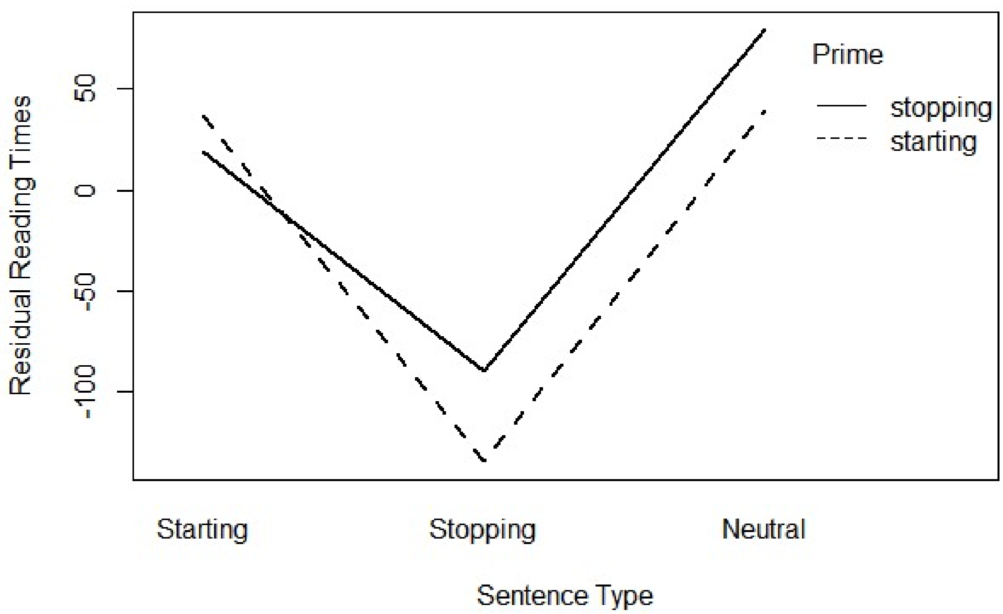


Figure 3.1 Critical Region Residual Reading Times by Condition

### 3.4.2 POST-CRITICAL REGION ANALYSIS

The post-critical region analysis followed the same steps as before. Post-critical region reading times were modelled for each participant using region character length as both a fixed and random factor (Trueswell et al., 1994). Then, a new model was fit to the residuals from the first model with sentence and prime type and their interaction as fixed factors and participants and items as random factors. With the Neutral sentence condition included, this model showed a significant effect of sentence type, such that the Neutral sentences were read more slowly than the Starting [$\beta = -826.2$, $t(27.1) = -3.582$, $p = .001$] or Stopping sentences [$\beta = -609.9$, $t(27.3) = -2.640$, $p = .014$]. With the Neutral condition excluded, the model showed no significant effects. The post-critical region residual reading data is shown in Figure 3.2.

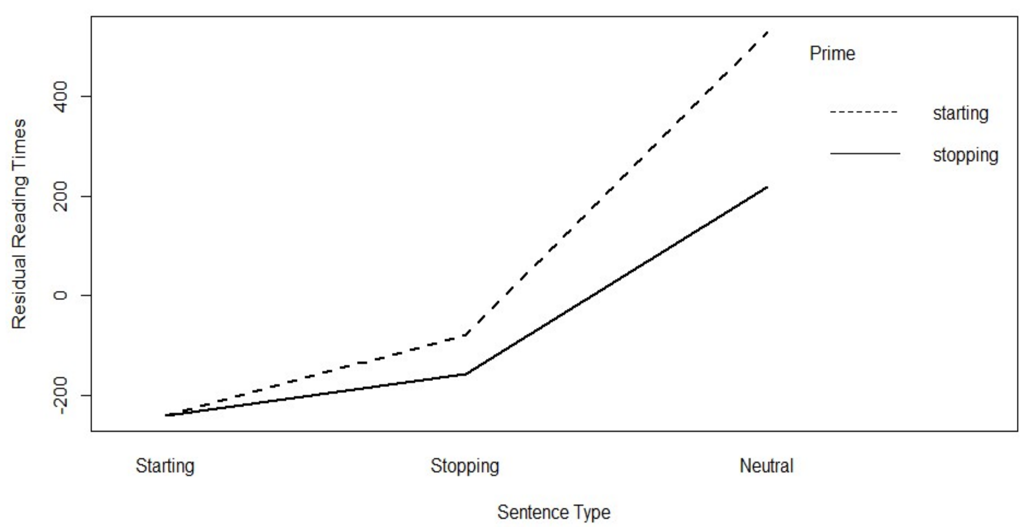


Figure 3.2 Post-Critical Region Residual Reading Times by Condition

### 3.4.3 POWER ANALYSIS

As in experiment one, in order to test whether the lack of an interaction between task and prime could be attributed to a lack of power, a post-hoc simulation-based power analysis was conducted for both the critical and post-critical regions using the simr package 1.0.5 in R 4.0.2 (Green & MacLeod, 2016). The power analysis followed exactly the same procedure as before. Assuming a relatively small effect size with a standardized $\beta = .15$, the simulations showed 85% power for the critical region and 91% power for the post-critical region. If the true effect size were even smaller at $\beta = .1$, the simulations showed only 62.5% power for the critical region and 60% power for the post-critical region.

## 3.5 DISCUSSION

The goal of experiment two was to test Talmy's (1988) further prediction that force dynamic image-schematic primes would facilitate the processing of abstract-metaphorical psychosocial causation sentences. As in experiment one, no interaction was found between prime and sentence type on the residual reading times either for the critical or the post-critical region. Therefore, this experiment does not provide evidence for the psychological reality of force dynamics. A significant effect was found such that the Neutral sentences were read more slowly than the Starting or Stopping sentences in the post-critical region. As in experiment one, this cannot be accounted for by word frequency, region length, as these factors were all either held constant between conditions or accounted for in the model. Further explanations for this effect will be considered in the general discussion.

# CHAPTER 4

## GENERAL DISCUSSION

The current study describes two experiments which were performed to test whether force dynamics (Talmy 1988) as a conceptual category affects the online processing of both concrete and abstract-metaphorical causal sentences. In order to test this hypothesis, participants completed interactive image-schematic primes which evoked either Starting or Stopping force dynamic schemas before reading matching, mismatching, or Neutral sentences. I predicted that, if Talmy's (1988) view was correct that I would find an interaction such that participants would read and validate the causal sentences more quickly when they matched the preceding prime. However, no such effect was observed either for the concrete or the abstract-metaphorical sentences.

In experiment one, the only significant effect found was that the Starting sentences were read more slowly than the Stopping and Neutral sentences. As noted in section 2.6, this effect cannot be attributed to syntax, word frequency, region word length, or region character length, as these were all either controlled between conditions (syntax and region word length) or included in the analysis (word frequency and region character length). Another possibility is that this difference is the result of differences in prototypicality between the schemas; however, as causation is more prototypical that prevention this again seems unlikely. Furthermore, as experiment two did not show the same effect, the schemas themselves cannot be the cause of this difference. The picture, admittedly, becomes more complicated when considering the thematic structure. None of the sentences in this experiment had an agentive subject, which may have been more incongruent with the more prototypical causation schema than the less prototypical prevention schema. The presence of animate patients in experiment two may have helped to bridge this gap, explaining the absence of this effect in experiment two. The remaining, rather less speculative, possibility is that this difference was due to surprisal differences which may have been triggered by subtle differences in collocational frequency and cloze probability between the conditions.

Experiment two showed a significant effect of sentence type such that the Neutral sentences were read more slowly than the Starting and Stopping sentences. Like the effect in experiment one, this cannot be attributed to word frequency, region word length, or region character length, for the reasons discussed above. An argument could be made that the causal primes encouraged participants to expect causal sentences, slowing the interpretation of noncausal Neutral sentences. However, this position is somewhat tenuous given the lack of any similar effect in experiment one, where there was no difference between the Neutral and Stopping sentences and the Neutral sentences were processed more quickly than the Starting sentences. However, in is this case, there is a small syntactic difference which may have caused the Neutral sentences to be more difficult to process. Specifically, in experiment two, the post-critical region involved a nonfinite verb phrase. For the critical sentences, the subject of this nonfinite verb was the object of the main verb, while for the neutral sentences, since they were not causal, the subject of the nonfinite verb was the subject of the main verb. So, for the Neutral sentences participants were required to go further back in the sentence to find the subject of the nonfinite verb, which may explain the slower reading times.

The most important result of these experiments was the lack of an interaction between sentence and prime type in both experiments one and two. There are several plausible explanations for failure of this study to find an effect which require further consideration—namely, 1) the strength, or lack thereof, of the manipulation, 2) statistical power, 3) the construction of the sentence stimuli, 4) the appropriateness of the self-paced reading paradigm, and 5) the existence of the effect itself.

Firstly, as to the strength of the manipulation, the goal of the interactive primes used in this study was to force participants to process the force dynamic schemas more deeply by participating in them rather than merely observing them (as was done in Madden and Pecher, 2010). One possibility is that these primes did not produce the depth of processing required in order to observe an effect. Since the primes themselves were easy to complete, it is possible that, over the course of the experiment, the primes became a mindless task which no longer required deep processing. If this were the case, I might expect an interaction of prime and sentence type with the order in which the trials were presented. However, no such effect was observed.[3]

In addition, it is unclear that such a strong manipulation should be required. Studies investigating other domains of grounded cognition have found robust effects using weaker manipulations. For example, Boroditsky (2000) and Boroditsky and

---

[3] To test this concern, I repeated the analyses for the critical and post-critical regions of both experiments including trial number as an additional fixed factor as well as the three-way interaction between trial number, sentence type, and prime type. For each analysis, the three-way interaction was not significant ($p > .05$)

Ramscar (2002) showed that image-schematic diagrams of spatial scenarios accompanied by validation questions were sufficient to prime sentence interpretation. Boroditsky and Ramscar (2002) further showed that watching a horse race was sufficient to prime grounded cognition about motion. Other studies have shown that merely hearing a hand related sentence can effect response times with the hand (Buccino et al., 2005) and that hearing a sentence implying a particular orientation for an everyday object can inhibit recognition of the same object in a different orientation (Stanfield and Zwaan, 2001).

Unlike in the current study, most of these primes were static rather than dynamic, and in none of them did participants have to directly interact with the schema being primed. Overall, while the manipulation in the current study could be strengthened (for example, by having participants interact with physical rather than virtual objects, a la Boroditsky and Ramscar's 2002 chair rodeo), it is far stronger than the manipulation used in previous studies on force dynamic grounding (namely as Madden and Pecher, 2010), in addition to being stronger than many successful manipulations in the wider grounding literature. Therefore, it is improbable that the lack of an effect observed here can be blamed primarily on the strength of the manipulation.

Secondly, as to statistical power, the power analyses conducted for each experiment showed that these experiments were sufficiently powered to find an effect, assuming a population effect size at standardized $\beta = .15$ (approximately equivalent to $R^2 = .0225$). If the population effect size were even smaller at standardized $\beta = .1$ (approximately $R^2 = .01$), then the experiments may not have had sufficient power. However, if the true effect size in the population is so small, it is unlikely to represent an important aspect of causal language processing. Overall, I argue that the experiments were sufficiently powered to find an effect of a meaningful size, if one exists in the population.

Thirdly, as for the construction of the sentence stimuli, there are two issues to consider, first, whether the sentences really evoked the force dynamic schemas they were intended to and second, what potential confounds the sentences introduced. As to the first point, while great care was taken to ensure that the stimuli sentences matched Talmy's (1988) schemas, none of the stimuli used in this experiment were unaltered examples from Talmy (1988). This was the case for several reasons. Firstly, in order to ensure equivalence between conditions, the sentences were all fit into the same simple syntactic frame. Talmy, in contrast, argued that in order to achieve maximum semantic simplicity, more complex syntax is sometimes required. Talmy (1988) compares the sentences in 13)

13) a. I broke the vase.

b. I broke the vase by hitting it with a ball.

c. The ball's hitting it broke the vase.

Talmy argues that while 13)a. might have the simplest syntax, 13)c. is semantically simpler as it represents a more basic precursor-event sequence. 13)a. in contrast requires a complex series of inferences, which are made explicit in 13)b., due in part to the presence of an agent. However, sentences like 13)c. would introduce additional confounds if used as stimuli sentence due to their highly marked syntax and complex event structure. In an attempt to sidestep this difficulty, the current experiment followed Talmy by using nonagentive sentences, but used a simpler syntactic construction as in 14).

14) The rubber ball knocked the heavy bowl off the table.

The use of an instrumental subject rather than a deverbal event subject results in significantly simpler syntax, which is more naturalistic and less likely to result in confusion for participants, while minimizing the number of inferences required to arrive at the basic precursor-event sequence, since the instrument is made explicit and both the precursor and consequent events are encoded in the verb.

Similarly, a great deal of care was taken to ensure that each sentence fit neatly within only one of Talmy's causal categories. For example, Talmy distinguishes between change-of-state and continuous causal paradigms. Therefore all of the stimuli were controlled to ensure that they could only be read as change-of-state events, to more closely match the events in the interactive primes. In addition, the sentences were controlled to ensure that the antagonist always appeared in subject position, resulting in more natural active sentences and mirroring the interactive stimuli in which participants always took the role of the antagonist. Overall, the stimuli closely matched Talmy's (1988) descriptions, while accounting for the syntactic constraints imposed by the experimental paradigm.

As for potential confounds introduced by the stimuli, as noted previously, semantic content, syntactic structure, and number of words were all matched across conditions. Furthermore, word frequency and character length were accounted for in the analysis. One possible remaining confound is the syntactic preferences of each causal schema. If it were the case that Starting and Stopping sentences had a different degree of preference for the syntactic structure used in these experiments, I would expect to see differences in reading times between the conditions, as was observed in experiment one in which the Starting sentences were read more slowly than the Stopping sentences.

However, the lack of a similar effect in experiment two suggests that this may not be the correct explanation. Further, even if this is the case, it would not explain the lack of an interaction between prime and sentence type in both experiments.

Fourthly, as to the self-paced reading paradigm, another possibility to be considered that self-paced reading is not an appropriate dependent variable to measure grounding effects. Self-paced reading paradigms are not prominent in the grounding literature, with most researchers preferring judgments (Boroditsky, 2000; Boroditsky & Ramscar, 2002; Gentner, Imai, & Boroditsky, 2002), action-sentence compatibility effects (Creem and Proffitt, 2001; Glover et al., 2004; Glenberg and Kaschak, 2002), or neurological measures (Aziz-Zadeh et al., 2006; Calvo-Merino et al., 2006). However, despite this, there are a number of extant studies showing grounding effects using self-paced reading. Wilson and Gibbs (2007) found that performing (or imagining performing) a physical activity (such as chewing) facilitated sentence reading times for compatible sentences. Similarly, Fekete (2012) found in one of his experiments that hearing sounds could facilitate or inhibit word-by-word self-paced reading times for compatible or incompatible sentences. And Vaci et al. (2014) was able to find a grounding effect in a region-by-region self-paced reading paradigm with sensibility judgments. Overall, while self-paced reading is not the most popular paradigm for examining grounding effects, these studies provide evidence that self-paced reading is a sensitive enough measure to uncover grounding effects when they are present.

The final possibility is that force dynamic image schemas do not influence the online comprehension of causal sentences. While it is dangerous to draw overly firm conclusions from null results, I argue that the balance of evidence uncovered in these experiments suggests that force dynamic schemas are not used in online processing. In two experiments, I have replicated Madden and Pecher's (2010) null result. And, given the statistical power and the strength of the manipulation involved in this study, it is likely that if such an effect exists, it is so small that it cannot contribute meaningfully to any theory of the processing of causal language.

Therefore, further research is required to show if or how causal language is grounded. One possibility is that the grounding of causal reasoning is based in more directly embodied physical experiences rather than in abstract image schemas, a possibility to which a psychological application of Talmy's (1988) force dynamic theory could be adapted. In other words, concepts such as CAUSING and PREVENTING may have the psychological significance that Talmy gives them, but if so, that significance is grounded in physical sensations such as pressure and muscle exertion rather than in image schemas. Further research should examine this possibility in more depth, both in reference to force dynamics and to competing theories like the focal set model (Wolff & Song, 2003).

## REFERENCES


Aziz-Zadeh, L., Wilson, M. S., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*(18).

Barr, D. J, Levy, R.,, Scheepers, C., & Tily, H. J. (2013). Random effects structure for confimatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3).

Boroditsky, L (2000). Metaphoric structuring: Understanding time through spacial metaphors. *Cognition, 75*(1), 1-28.

Boroditsky, L & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*(2), 185-189.

Brookes, J., Warburton, M., Alghadier, M, Mon-Williams, M. & Muchtaq, F. (2019) Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavioral Research Methods, 52*, 455-463.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research, 24*(3), 355-363.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology 16*(19).

Cheng, P. W. & Novick, L. R. (1992). Covariation in Natural Causal Induction. *Psychological Review, 99*(2).

Creem, S. H. & Proffitt, D. R. (2001). Grasping objects by their handles: A necessary interaction between cognition and action. *Journal of Experimental Psychology, 27*(1), 218-228.

Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA).* Available online at https://www.english-corpora.org/coca/.

Evans, V. & Green, M. (2006). *Cognitive linguistics: An introduction.* Lawrence Eribaum Associates Publishers.

Fauconnier, G. & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*, 133-187.

Fekete, István. (2012). A corpus- and psycholinguistic testing of the strong version of the embodiment hypothosis in cognitive linguistics. [Doctoral dissertation, Budapest University of Technology and Economics]

Fischer, M. H. & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology, 61*(6), 825-850.

Gentner, D., Imai, M. & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space → time metaphors. *Language and Cognitive Processes, 17*(5), 537-565.

Gibbs, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind & Language, 21*(3), 434-458.

Gibbs, R. W. (2008). Metaphor and thought: The state of the art. In R. W. Gibbs, Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (p. 3–13). Cambridge University Press.

Gibbs, R. W., Gould, J.J. & Andric, M. (2006). Imagining metaphorical actions: Embodied simulations make the impossible plausible. *Imagination, Cognition and Personality, 25*(3), 221-238.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D. & Buccino, G. (2008). Processing abstract language modulates motor system activity. *The Quarterly Journal of Experimental Psychology, 61*(6), 905-919.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*, 103-108.

Green, P. & MacLeod, C.J. (2016). Simr: An R package for power analysis of generalised mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498.

Hubert, Mia & Vandervieren, Ellen. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis, 52*.

Jamrozik, A., McQuire, M., Cardillo, E. R., & Chaterjee, A. (2015). Metaphor: Bridging embodiment to abstractions. *Psychonomics Bulletin & Review, 23*(4).

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. doi: 10.18637/jss.v082.i13

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* The University of Chicago Press.

Madden, C. & Pecher, D. (2010). The force behind language: Are concrete and abstract sentences understood in terms of underlying force patterns? Unpublished manuscript.

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. Ross (Ed.). The Psychology of Learning and Motivation, vol. 54 (pp. 217-248). Burlington: Academic Press.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna, Austria.

Stanfield, R. A. & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2).

Talmy, L. (1988) Force dynamics in language and cognition. *Cognitive Science, 12*, 49-100.

Thibodeau, P. H. & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS One.*

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 1994, 285-318.

Vaci, N., Radanović, J., Marmolejo-Ramos, F., & Milin P. (2014). Comprehending negated action(s): Embodiment perspective. *Cognitive Processing, 15* (Supplement 1)

Wilson, N. L. & Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science, 31*, 721-731.

Wolff, P. & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*(3), 276-332.

## APPENDIX A

## EXPERIMENT ONE STIMULI SENTENCES

Table A.1 Critical sentences used in Experiment 1.

| Starting | Stopping |
|---|---|
| The rushing flood washed the sand bags into the house. | The sand bags blocked the rushing flood from the house. |
| The floating debris dragged the fishing nets down the river. | The fishing nets snagged the floating debris in the river. |
| The draining water pulled the hair clog down the pipe. | The sudden clog trapped the draining water in the pipe. |
| The race car launched the flimsy barrier off the track. | The heavy barrier blockaded the race car on the track. |
| The flying rock spread the windshield fragments around the car. | The thick windshield stopped the flying rock outside the car. |
| The fast flyball shoved the safety net toward the stands. | The safety net caught the fast flyball in the stadium. |
| The golf ball pushed the light grass into the hole. | The thick grass entangled the golf ball outside the hole. |
| The speeding bullet drove the kevlar shred into the liver. | The kevlar fabric ensnared the speeding bullet in the vest. |
| The rubber ball knocked the heavy bowl off the table. | The heavy bowl stopped the rubber ball on the table. |
| The rolling log forced the thick mud over the ledge. | The thick mud captured the rolling log near the ledge. |

Table A.2 Nonsense and Neutral sentences used in Experiment 1.

| Nonsense | Neutral |
|---|---|
| The gray house plugged the bath water in the tub. | The swift stream followed the narrow canyon to the sea. |
| The ancient building yanked the soccer ball into the road. | The rising tides approached the docks near the shore. |
| The garbage can grabbed the car keys from the hook. | The flash flood reached the new buildings on the hill. |
| The white fence drank the red wine from the glass. | The green car crossed the busy street near the park. |
| The fire hydrant ate the juicy apple on the table. | The loud motorcycle passed the old building outside the university. |
| The thin curtains rammed the fire truck into the building. | The bowling ball missed the white pins in the alley. |
| The mighty wave silenced the glaring sun on the beach. | The soaring baseball exited the playing field into the stands. |
| The new ship halted the train cars on the hill. | The stray arrow overshot the small target in the field. |
| The loud dishwasher prevented the powerful ship in its journey. | The rolling ball neared the cat bed in the den. |
| The blue boat tugged the potted plants from the shop. | The light breeze entered the quiet bedroom through the window. |
| The old tree shook the black microwaves from the branches. | |
| The sink plug stemmed the gushing river in the mountains. | |
| The red truck navigated the file folders in the cabinet. | |
| The key ring suspended the rolling stones from the path. | |
| The swift current floated the small canoe to the moon. | |
| The old cannon fired the lead ball across the galaxy. | |
| The steel cable towed the damaged car into the lamp. | |
| The salty ocean dropped the little boat into the soup. | |
| The strong wind impeded the pink umbrella from the dishwasher. | |
| The dangerous flood swept the old cars under the rug. | |

APPENDIX B

EXPERIMENT TWO STIMULI SENTENCES

Table B.1 Critical sentences used in Experiment 2

| Starting | Stopping |
|---|---|
| The brilliant writing dragged the reluctant schoolboy into reading the novel. | The impenetrable writing barred the eager schoolboy from reading the novel. |
| The inspiring example launched the hesitant entrepreneur into starting the business. | The bureaucratic hurdles blockaded the hopeful entrepreneur from starting a business. |
| The persistent chiding pushed the indecisive artist into finishing the painting. | The constant distractions blocked the diligent artist from finishing the painting. |
| The looming deadline pressured the procrastinating student into completing the essay. | The distracting drama restrained the dedicated student from completing the essay. |
| The sudden opportunity drove the cautious businessman into taking the risk. | The sudden doubts stopped the bold businessman from taking the risk. |
| The strong evidence spurred the lenient judge into increasing the sentence. | The weak evidence obstructed the pitiless judge from increasing the sentence. |
| The public outcry moved the dishonest politician into keeping the promise. | The public outcry held the honest politician from keeping the promise. |
| The worsening symptoms pressed the defiant patient into accepting the treatment. | The staggering costs restricted the critical patient from accepting the treatment. |
| The enticing advertisements pulled the uncertain customer into buying the car. | The financial crisis impeded the excited customer from buying the car. |
| The rapturous applause propelled the timid musician into playing the encore. | The lackluster applause snagged the enthusiastic musician from playing the encore. |

Table B.2 Nonsense and Neutral sentences used in Experiment 2.

| Nonsense | Neutral |
|---|---|
| The intoxicating pleasure drank the party goer without eating the cake. | The bad luck followed the careful gambler without ruining his income. |
| The uncomfortable rejection thought the young banker from liking the changes. | The speeding bullet missed the brazen soldier without impairing his confidence. |
| The wonderful serendipity trusted the beautiful adventurer into reaching the moon. | The flash flood reached the new homeowners without damaging their home. |
| The irritating foolishness flocked the conditioned athlete into being the bus. | The loud motorcycle passed the old pedestrian without hurting his hearing. |
| The warm greeting ate the artistic baker without noticing him. | The stray arrow overshot the oblivious deer without making a sound. |
| The restrictive knitting relapsed the recovering alcoholic into singing the Venice. | The bouncing ball neared the sleeping cat without creating a stir. |
| The spilled milk resurrected the transformational speaker into driving the car. | The heavy clouds left the partying beachgoers without starting a storm. |
| The statistical sophistication read the rejoicing preacher from growing a plant. | The concerned public petitioned the apathetic dictator without changing his mind. |
| The delicate carving diversified the sales team from working the weekends. | The frequent inconveniences accompanied the sturdy mechanic without affecting his mood. |
| The playful rabbit disintegrated the crazed warlord without reaching the waterfall. | The occasional worry visited the speculative investor without altering his strategy. |
| The military intelligence concerned the careful general into exploding the bookcase. | |
| The spreading jealousy saddened the humble minister from flapping his nose. | |
| The cityscape's beauty dazzled the excited tourist without forging a boat. | |
| The winter coldness impressed the freezing southerner without rowing the boat. | |
| The army recruiter enlisted the high schooler into grounding the mountain. | |
| The critical newspaper reviewed the concert pianist from rebounding the baseball. | |

| Nonsense | Neutral |
| --- | --- |
| The unexpected consequences chastised the unprepared child into eating the moon. | |
| The prestigious award honored the brilliant chemist from indulging the aliens. | |
| The important experience taught the novice boxer from petrifying the cake. | |
| The horrible smell repelled the food critic without awakening the streets. | |

APPENDIX C

ADDITIONAL ANALYSES

In addition to residual reading times, both raw reading times and reading rate (reading time divided by region character length) were tested as dependent variables. These analyses did not reveal any effects which were not found in the residual reading time analysis. For each of these models, log word frequencies from the Corpus of Contemporary American English (Davies, 2008-) were included as random effect for each of the content words in the critical region. However, this inclusion did not improve the models.

C.1 EXPERIMENT ONE CRITICAL REGION ANALYSIS

Three additional models were tested for the critical region in experiment one. The first two models used the raw region reading times as a dependent measure. In the first model, the reading time for the critical region was modeled based on the interaction between sentence type and prime type as well as region character length as fixed factors, with participants and item number as random factors. With the Neutral condition included as a reference, this model showed significant main effects of sentence type, such that the Starting sentences were read more slowly than the Neutral sentences [$\beta = 253.4$, $t(27.4) = 2.672$, $p = .013$], and of region character length, such that longer regions were read more slowly [$\beta = 31.33$, $t(24.9) = 2.698$, $p = .012$]. However, there were no significant interactions. Excluding the Neutral condition revealed a significant effect of sentence type, such that the Starting sentences were read more slowly than the Stopping sentences [$\beta$ = -267.2, t(22.6) = -3.303, p = .003] and of region character length such that longer items were read more slowly than shorter items [$\beta$ = 47.42, t(17) = 3.702, p = .002]. A second model was fit to the data which differed from the first only in treating region character length as random, rather than a fixed, factor. With the Neutral condition included as a reference, the model showed the same significant effect of sentence type, such that the Starting sentences were read more slowly than the Neutral sentences [$\beta$ = 210.4, t(25.8) = 2.593, p = .015]. When the Neutral condition was excluded, there were no significant effects.

The third model used log reading rate (critical region reading time divided by region character length) as a dependent measure, modeling it based on the interaction of sentence and prime type, including participants and item number as random effects. With the Neutral condition included, the final model showed no significant effects from or interactions between any of the fixed factors ($p > .05$), though there was a trend towards significance for the same effect of sentence type as noted in the previous two models (p = .066), however, this model was not a significant improvement over a null model which included only the random effects [$\chi2(5)$ = 12.619, p = .027]. Excluding the Neutral condition, the model showed no significant effects, and was not a significant improvement over the null model [$\chi2(3)$ = 7.060, p = .070].

C.2 EXPERIMENT ONE POST-CRITICAL REGION ANALYSIS

As in the critical region, three additional models were tested for the post-critical region. In the first model, the raw reading time for the post-critical region was modeled based on the interaction of sentence type and prime type as well as region character length as fixed factors, while including participants and item number as random factors.

This model showed a significant effect of sentence type such that the Starting sentences were read more slowly than the Neutral sentences [$\beta = 319.2$, $t(24.4) = 2.159$, $p = .041$]. Excluding the Neutral condition, the model showed no significant effects.

A second model, only differing from in first in treating region character length as a random effect, was fit to the data. With the Neutral condition included, this model showed the same effect of sentence type [$t(17.8) = 3.383$, $p = .034$)]. When the Neutral condition was excluded, the model this model now showed a significant main effect of sentence type such that the Stopping sentences were read more quickly than the Starting sentences [$\beta = -344.5$, $t(15) = -2.57$, $p = .021$].

A third model was fit using log reading rate (post-critical region reading time divided by region character length) as the dependent variable. This model used the interaction of sentence type and prime type as fixed factors, and participant and item number as random factors. With the Neutral condition included, the model showed no significant effects or improvement over a null model ($p > .05$). When the Neutral condition was excluded, the model showed a significant effect of prime type such that the post-critical regions were read more slowly after the Starting primes [$\beta = -.06$, $t(1103) = 2.043$, $p = .041$] and a near significant interaction between prime and sentence type, such that the post critical region was read more slowly when sentence type matched prime type [$\beta = .07$, $t(1103) = 1.916$, $p = .056$]. However, once again, this model was not a significant improvement over the null model ($p > .05$).

## C.3 EXPERIMENT TWO CRITICAL REGION ANALYSIS

The same additional analyses were performed for experiment two as for experiment one. The first model predicted the reading time of the critical region using the interaction of sentence type and prime as well as region character length as fixed factors, while including participant and item number as random factors. With the Neutral condition included, this model showed only a main effect of region character length [$\beta$ = 41.2, $t(24.6) = 2.974$, $p = .006$]. With the Neutral condition excluded, the model showed no significant effects. Adding log word frequencies did significantly improve this model both with and without the Neutral sentence condition [$\chi2(3) = 12.493$, $p = .006$], showing a significant main effect of verb frequency such that sentences with more frequent verbs were read more quickly [$\beta = -49.9$, $t(21.9) = -2.103$, $p = .047$]. This was the only case in which including word frequencies improved one of the models.

The second model was identical to the first except that it treated region character length as a random factor. This change did not reveal any significant effects ($p > .05$).

Finally, the log reading rate of the critical region (region reading time divided by region character length) was modelled based on the interaction of prime and sentence type as fixed factors with participant and item as random factors. Both with and without the Neutral condition included, this model showed no significant effects or improvement compared to a null model ($p > .05$).

## C.4 EXPERIMENT TWO POST-CRITICAL REGION ANALYSIS

The additional analyses for the post-critical region followed the same procedure as before. First, the post-critical region reading time was modelled using the interaction of sentence type and prime as well as region character length as fixed factors, with participant and item as random factors. With the Neutral condition included, the model showed a main effect of sentence type such that the Neutral sentences were read more slowly than the Starting sentences [$\beta = -707.3$, $t(26) = -3.355$, $p = .002$] and Stopping sentences [$\beta = -519.6$, $t(26.5) = -2.452$, $p = .021$]. When the Neutral condition was excluded, this model showed no significant effects.

The second model only differed from the first in treating region character length as a random rather than a fixed factor. With the Neutral condition included, the model showed the same significant effect of sentence type, such that the Neutral sentences were read more slowly than the Starting sentences [$\beta = -763.2$, $t(12.5) = -3.609$, $p = .003$] and Stopping sentences [$\beta = -571.2$, $t(13.5) = -2.688$ , $p = .018$]. Without the Neutral condition, the model showed no significant effects ($p > .05$).

Finally, the log reading rate of the post-critical region (region reading time divided by region character length) was modelled based on the interaction of sentence and prime type as fixed factors with participant and item as random factors. With the Neutral sentences included, this model again showed a significant effect of task such that the Starting [$\beta = -.33$, $t(26.82) = -3.363$, $p = .002$] and Stopping [$\beta = -.25$, $t(26.91) = -2.521$, $p = .018$] sentences were read more quickly than the Neutral sentences. However, this model was not a significant improvement over the null model (p . .05). When the Neutral condition the model showed no significant effects, and no improvement over a null model ($p > .05$).