Title
Space and Time: Relationships among Language, Body, and Mind

Permalink
https://escholarship.org/uc/item/2mk948b0

Author
Hendricks, Rose Katherine

Publication Date
2017

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Space and Time: Relationships among Language, Body, and Mind

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Cognitive Science by Rose Katherine Hendricks

Committee in charge:
Professor Lera Boroditsky, Chair
Professor Benjamin K. Bergen, Co-Chair
Professor David Barner
Professor Seana Coulson
Professor Federico Rossano

2017
The dissertation of Rose Katherine Hendricks is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2017
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page ........................................ iii</td>
</tr>
<tr>
<td>Table of Contents ....................................... iv</td>
</tr>
<tr>
<td>List of Figures ........................................ ix</td>
</tr>
<tr>
<td>List of Tables ........................................... x</td>
</tr>
<tr>
<td>Acknowledgements ....................................... xi</td>
</tr>
<tr>
<td>Vita .................................................... xii</td>
</tr>
<tr>
<td>Abstract of the Dissertation ........................... xiii</td>
</tr>
<tr>
<td>Chapter 1</td>
</tr>
<tr>
<td>1.1 Metaphor &amp; Thought .................................. 1</td>
</tr>
<tr>
<td>1.2 Language &amp; Thought .................................. 3</td>
</tr>
<tr>
<td>1.2.1 The Causal Problem ............................... 6</td>
</tr>
<tr>
<td>1.2.2 The Nature of Effects of Language on Thought .. 9</td>
</tr>
<tr>
<td>1.3 Spatio-Temporal Metaphors: Bridging Case Study .... 11</td>
</tr>
<tr>
<td>1.3.1 Does Language Play a Causal Role? ................ 15</td>
</tr>
<tr>
<td>1.3.2 Mechanisms: Testing Verbal &amp; Spatial Working Memory with Dual Tasks ......... 17</td>
</tr>
<tr>
<td>1.4 Overview of Studies ................................ 19</td>
</tr>
<tr>
<td>Chapter 2</td>
</tr>
<tr>
<td>2.1 Abstract ............................................. 23</td>
</tr>
<tr>
<td>2.2 Experiment 1 Introduction ............................ 24</td>
</tr>
<tr>
<td>2.2.1 Spatio-Temporal Metaphors: A Brief Overview .... 24</td>
</tr>
<tr>
<td>2.2.2 Experiment 1 Overview ............................. 26</td>
</tr>
<tr>
<td>2.3 Experiment 1 Methods ................................ 27</td>
</tr>
<tr>
<td>2.3.1 Participants ....................................... 27</td>
</tr>
<tr>
<td>2.3.2 Procedure .......................................... 28</td>
</tr>
<tr>
<td>2.3.3 Participant &amp; Trial Inclusion Criteria ........... 32</td>
</tr>
<tr>
<td>2.3.4 Analyses ........................................... 33</td>
</tr>
<tr>
<td>2.4 Experiment 1 Results ................................ 34</td>
</tr>
<tr>
<td>2.4.1 Main (Hypothesis-Relevant) Results ............... 34</td>
</tr>
<tr>
<td>2.4.2 Ancillary Results .................................. 35</td>
</tr>
<tr>
<td>2.5 Experiment 1 Discussion ............................. 38</td>
</tr>
<tr>
<td>2.5.1 Open Questions ..................................... 40</td>
</tr>
<tr>
<td>2.6 Experiment 2 Introduction ........................... 41</td>
</tr>
</tbody>
</table>
Chapter 3 Chinese speakers’ space-time associations

3.1 Abstract .......................... 47
3.2 Introduction ........................ 48
3.3 Experiment 3 Methods ......... 52
  3.3.1 Participants & Trial Inclusion ..... 52
  3.3.2 Materials & Procedure .......... 53
  3.3.3 Analyses ........................ 54
3.4 Experiment 3 Results .......... 55
  3.4.1 Time Judgment Reaction Times 55
  3.4.2 Time Judgment Accuracies .... 59
  3.4.3 Interference Accuracies ....... 59
  3.4.4 Comparison to Experiment 1, Metaphor Training 61
3.5 Experiment 3 Discussion ...... 62
3.6 Experiment 4 Methods .......... 63
  3.6.1 Participants & Trial Inclusion .. 63
  3.6.2 Materials & Procedure .......... 63
3.7 Experiment 4 Results .......... 64
  3.7.1 Time Judgment Reaction Times 64
  3.7.2 Proficiency & Congruency Effects 65
  3.7.3 Accuracies .................... 66
  3.7.4 Comparison to Experiment 3, Natural Language Congruency Effects 68
  3.7.5 Comparison to Experiment 1, Metaphor Training for English Speakers 68
3.8 Experiment 4 Discussion ....... 70
3.9 General Discussion .............. 70

Chapter 4 Consistencies in space-time and space-number associations under spatial interference

4.1 Abstract .......................... 74
4.2 Introduction ........................ 74
4.3 Experiment 5: Space-Time Associations from Visuomotor Experiences .... 78
4.4 Experiment 5 Methods .......... 78
<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Interference Calibration</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Task with Interference</td>
<td>32</td>
</tr>
<tr>
<td>2.3</td>
<td>Results of Experiment 1</td>
<td>36</td>
</tr>
<tr>
<td>2.4</td>
<td>Results of Experiment 2</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>Results of Experiment 3</td>
<td>56</td>
</tr>
<tr>
<td>3.2</td>
<td>Results of Experiment 4</td>
<td>66</td>
</tr>
<tr>
<td>4.1</td>
<td>Results of Experiment 5</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>Results for Experiment 5 vs. Experiment 1</td>
<td>83</td>
</tr>
<tr>
<td>4.3</td>
<td>Results of Experiments 6 &amp; 7</td>
<td>91</td>
</tr>
<tr>
<td>4.4</td>
<td>Results of Experiments 8</td>
<td>100</td>
</tr>
<tr>
<td>5.1</td>
<td>Congruency effects by experiment and interference type</td>
<td>106</td>
</tr>
<tr>
<td>5.2</td>
<td>Congruency Effects over 56 Trials</td>
<td>108</td>
</tr>
<tr>
<td>5.3</td>
<td>Congruency effects by experiment and block order.</td>
<td>110</td>
</tr>
<tr>
<td>5.4</td>
<td>Earlier-is-up biases</td>
<td>113</td>
</tr>
<tr>
<td>5.5</td>
<td>Congruency effects by stimulus duration</td>
<td>115</td>
</tr>
<tr>
<td>5.6</td>
<td>Congruency effects by number of interference distractors</td>
<td>116</td>
</tr>
<tr>
<td>6.1</td>
<td>Gesture for &quot;tomorrow&quot;</td>
<td>129</td>
</tr>
<tr>
<td>6.2</td>
<td>Gesture for &quot;earlier&quot;</td>
<td>130</td>
</tr>
<tr>
<td>6.3</td>
<td>Results of Experiment 9</td>
<td>131</td>
</tr>
<tr>
<td>6.4</td>
<td>Experiment 9 by items</td>
<td>132</td>
</tr>
<tr>
<td>7.1</td>
<td>Results of Study 10</td>
<td>150</td>
</tr>
<tr>
<td>7.2</td>
<td>Results of Study 11</td>
<td>159</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1: Results of Experiment 1: Reaction Times . . . . . . . . . . . . . . 35
Table 2.2: Experiment 1 Time Judgment Accuracies (%) . . . . . . . . . . . . 37
Table 2.3: Experiment 1 Interference Memory Accuracies (%) . . . . . . . . 37

Table 3.1: Experiment 3 Reaction Times . . . . . . . . . . . . . . . . . . . . 56
Table 3.2: Experiment 3 Time Judgment Accuracies (%) . . . . . . . . . . . 60
Table 3.3: Experiment 3 Interference Accuracies (%) . . . . . . . . . . . . . 61
Table 3.4: Experiment 4 Time Judgment Reaction Times . . . . . . . . . . . 65
Table 3.5: Experiment 4 Time Judgment Accuracies (%) . . . . . . . . . . . 67
Table 3.6: Experiment 4 Interference Accuracies (%) . . . . . . . . . . . . . 67

Table 4.1: Results of Experiment 5 . . . . . . . . . . . . . . . . . . . . . . . . 79
Table 4.2: Results for Experiments 6 & 7 . . . . . . . . . . . . . . . . . . . 91
Table 4.3: Judgment accuracies for Experiments 6 and 7 (%) . . . . . . . . . 92
Table 4.4: Experiments 6 and 7 Interference Accuracies (%) . . . . . . . . . 93

Table 7.1: Study 11: Z-scored acceptability ratings . . . . . . . . . . . . . . . 160
Table 7.2: Performance on the forced choice interpretation task . . . . . . . 160
ACKNOWLEDGEMENTS

I sincerely thank my committee for helpful feedback on the work presented here. I also thank co-authors Esther Walker and Tyler Marghetis, who contributed greatly to studies in this dissertation and served as invaluable mentors.

Chapter 2, in part, is a reprint of the material as it appears in Hendricks, R. K. & Boroditsky, L. (2017). New Space-Time Metaphors Foster New Nonlinguistic Representations, *Topics in Cognitive Science*. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the material as it appears in Hendricks, R. K. & Boroditsky, L. (2017). New Space-Time Metaphors Foster New Nonlinguistic Representations, *Topics in Cognitive Science*. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Hendricks, R. K., Walker, E. J., Bergen, B. K. & Boroditsky, L. (2017). Consistencies in space-time and space-number associations under spatial interference. *Acta Psychologica*. The dissertation author was the primary investigator and author of this paper.

Chapter 7, in part, has been submitted for publication of the material as it may appear in Hendricks, R. K., Bergen, B. K. & Marghetis, T. (2017). Do metaphors move from mind to mouth? Evidence from a new system of linguistic metaphors for time. *Cognitive Science*. The dissertation author was the primary investigator and author of this paper.

This work was also supported by an NSF GRFP grant, DGE-1650112.
VITA

2013 B. S. in Cognitive Science, Vassar College
2017 Ph. D. in Cognitive Science, University of California, San Diego

PUBLICATIONS


FIELDS OF STUDY

Major Field: Cognitive Science
ABSTRACT OF THE DISSERTATION

Space and Time: Relationships among Language, Body, and Mind

by

Rose Katherine Hendricks

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2017

Professor Lera Boroditsky, Chair
Professor Benjamin K. Bergen, Co-Chair

How do we construct our knowledge of abstract concepts like justice, free-will, or time that we can’t physically experience? To what extent can language, and in particular, linguistic metaphor, shape our mental representations? These are core questions for two prominent lines of work in Cognitive Science – one focused on metaphor, and the other on the relationship between language and thought. The experiments I’ll discuss use the domain of time to bridge these two lines of work. The studies shed light on metaphor’s capacity to create nonlinguistic representations that are flexible and dynamic. The relationship between linguistic metaphor and mental representation is bidirectional:

xiii
metaphors in language can create new ways of thinking, and preexisting patterns in thought can give rise to new linguistic metaphors.
Chapter 1

Introduction

1.1 Metaphor & Thought

Humans have the unique ability to conceptualize and reason about concepts we can neither see nor touch. For example, we have extensive memories of the past and can foresee and plan for the future. Some scholars believe these abilities have likely provided an evolutionary advantage that sets us apart from other animals [111]. But how do we construct these concepts?

One possibility is that we don’t construct this knowledge at all; instead, these intangible and invisible concepts may be innate. Chomsky [24], for example, based this claim on the belief that our everyday experiences, in language and otherwise, are not rich enough to possibly create a basis for the complex ideas we hold. The only explanation for how we can hold such rich mental representations given the poverty of the stimuli we receive, then, is that the concepts are innate. This is a primarily philosophical account, and is difficult to falsify experimentally.

Another account focuses on the role of our lived experiences, suggesting that when we consistently experience multiple sensations together, we associate those concepts [74].
For example, as infants we’re likely to experience physical warmth when were receiving affection from a caregiver, leading to the mental association, or conceptual metaphor, that AFFECTION is WARMTH. Similarly, as we move through space, we constantly experience time progressing simultaneously, giving rise to the conceptual metaphor that TIME is SPACE.

Initially, scholars primarily invoked linguistic evidence to suggest that we hold cross-domain associations between abstract and more concrete concepts. [91, 92] For example, Lakoff and Johnson [74] pointed out that we use many families of metaphors that draw from a single source domain to talk about abstract ideas: We say things like I’m feeling low, down in the dumps, or things are looking up, which Lakoff and Johnson presumed to reflect a mental association between vertical space and happiness, often referred to as a conceptual metaphor. However, linguistic evidence alone cannot shed light on how these mental concepts come to be [91, 92]. Instead, linguistic evidence gives rise to a chicken-and-egg problem: Do systematic metaphors in language primarily reflect pre-existing mental cross-domain associations, or do the metaphors in language give rise to those mental associations? Perhaps both forces are at play.

Following from the inconclusive nature of linguistic evidence, a third account suggests that metaphors in language may create cross-domain associations in mind that support our construction of rich mental representations of concepts like love, justice, or time that we can’t physically experience. A rich line of work on the psychological consequences of linguistic metaphors has arisen from this account, showing that, at least in the short term, metaphors can shape behavior in systematic ways by guiding the knowledge we activate (for review, see [115]).

For instance, in one study, Alan Turing was seen as more of a genius with more exceptional inventions when his ideas were described as light bulbs rather than seeds [39]. In other work, people were more likely to support reform, rather than enforcement-
oriented, approaches to crime reduction when crime was described as a virus than as a beast [113, 114, 26]. Other studies have shown that personifying changes in stock prices (climbing and slipping), rather than objectifying them (increasing and decreasing in value), makes people more likely to think recent price trajectories will continue into the future [89, 69]. And framing cancer as an enemy in a war has been found to reduce people’s intentions to engage in self-limiting preventative behaviors (e.g., eating less red meat, smoking less [62]; and to think that it would be harder for cancer patients to come to terms with their situation [64].

Thus, linguistic metaphors can shape thought in the moment by highlighting some features of our mental representations and masking others. But can linguistic metaphor be involved in the creation of our representations? And if so, what is the mechanism through which metaphors in language create new ways of thinking? Does metaphor’s influence remain in the linguistic sphere, or can metaphorical language foster nonlinguistic mental representations?

The work presented in this dissertation addresses these questions in the context of driving questions from research on the relationship between language and thought.

1.2 Language & Thought

The broad question Does language shape thought? is made up of numerous sub-questions: Do speakers of different languages think about the world differently? Does learning a new language cause a person to think in a new way? Are there some thoughts that are unthinkable without language, or without knowing a specific language? Scholars from diverse academic backgrounds with varied assumptions have approached these questions using a range of methods. As a result, accounts of the relationship between language and thought are correspondingly diverse. In the work presented here, I
focus specifically on whether patterns in language can systematically guide speakers to construct mental representations.

One view is that language and thought are independent [99]. According to this philosophical view, we think in a language-like system, referred to as mentalese, which establishes the concepts that linguistic words are mapped to. Because of the supposed independence of language and thought, then, the possibility that language may shape thought is rendered illogical.

An alternative account grants that language and thought may be connected, but that language does not shape thought; instead, language is a byproduct of pre-existing thoughts (i.e., [23]). Some empirical evidence has been interpreted as consistent with this account. For example, Heider and Oliver [63] showed that a preliterate tribe from New Guinea, the Dugum Dani, who use only two linguistic labels for colors, remembered and categorized new hues in a way that was comparable to how English speakers did. Gleitman and Papafragou [53] interpret evidence like this study as demonstrating that thought comes first, and language is its expression.

These two accounts contrast with a third, often referred to as linguistic relativity, which posits that the structure of language influences the way that speakers perceive and understand the world [48]. A large portion of linguistic relativity research relies on cross-linguistic methods, comparing how people who speak languages that differ in some critical feature perform on a related task.

For example, speakers of Guugu Yimithirr, an Australian aboriginal language, rely on an absolute frame of reference to talk about space. That is, they describe locations in terms of cardinal directions (north, south, east, and west) as opposed to relative relations (left and right), as speakers of Western languages tend to do. Consistent with these linguistic differences, when participants were asked to remember locations of objects or places, Guugu Yimithirr speakers recalled them with respect to their absolute locations,
while Dutch speakers did so with respect to their relative locations [76].

Extensive research has also been conducted on linguistic relativity using the domain of color. A number of studies have shown that people demonstrate categorical perception, distinguishing hues that cross a linguistic category boundary more quickly than hues from the same linguistic color category (for review, see [101]). For instance, different hues that fall under the label *blue* in English obligatorily fall under two separate labels in Russian: *siniy* and *goluboy*. Prior work has shown that Russian speakers are faster to complete a color-matching task when the two hues they must distinguish between differ in their linguistic labels – when one is siniy and the other is goluboy [127]. English speakers do not show the same facilitation when the two shades are on opposite sides of the Russian blue boundary, suggesting that because Russian requires its speakers to habitually attend to the siniy/goluboy distinction in their environment (in order to express it in language), they are physically faster to differentiate these hues from each other.

Additional work has shown that patterns in language can affect memory for features of events like the people involved [43] and their progression [2], as well as attention to motion [97]. It can alter the way people conceptualize objects, for example according to their grammatical gender [105] or their spatial relationship – specifically, the tightness of fit between two objects [22], or the material they’re made of versus their function [79].

In other cases, research has shown that linguistic and nonlinguistic representations can work together, allowing people to conceptualize the world in a way they wouldn’t be able to with either type of representation alone. Wolff and Holmes [129] call this phenomenon *language as augmenter*. For example, language seems to be necessary for developing the ability to count exact numbers. The Piraha, who lack exact quantity vocabulary, resort to an approximate number system when completing tasks that require the use of an exact number system [56, 42]. Within a single culture, Spaepen and
colleagues [109] showed that deaf Nicaraguans who have not learned a signed language perform similarly to the Piraha on tasks that require them to count exact quantities, even though they live in a numerate culture.

Across a variety of domains, the wealth of cross-cultural work on language and thought demonstrates robust correspondences between patterns in language and patterns in thought. However, this body of work leaves two important questions unanswered: First, can patterns in language actually cause people to think in different ways? And if so, can language create patterns in thought that are nonlinguistic? I discuss these questions in the following two sections.

1.2.1 The Causal Problem

Documenting correlations between patterns in language and patterns in thought cannot establish that language plays a causal role in how we construct mental representations. Whenever studies compare behavior across different linguistic groups, the design is necessarily quasi-experimental. It is not possible to randomly assign participants to be native speakers of any language - they are already native speakers of a particular language when they begin the experiment, and with differences in native languages come a potentially infinite set of other confounding factors.

For this reason, some researchers have proposed that cross-linguistic language and thought research supports the inverse of linguistic relativity. According to this view, patterns in language do not shape thought; instead, underlying patterns in thought, which differ from one culture to the next, shape our language [53]. This view is reflected, for example, in Chomsky’s [23] assertion that by studying the ways that languages differ, we can gain insight into cognitive differences across cultures.

Another possible explanation for research that demonstrates correspondences between language and thought is that these correspondences arise because of some
extraneous variable, a set of underlying cultural differences unrelated to language. For example, some hypothetical feature of Russian culture, such as the pervasiveness of blues in everyday objects or landscapes, may make it more important to differentiate shades of blue than in American culture. A cultural emphasis on blues could, in turn, have led to the establishment of the Russian linguistic distinction between siniy and goluboy, and the cultural emphasis may have also trained Russians to carefully discern shades of blue, a skill demonstrated by their facilitation for cross-boundary color discriminations in prior work [127].

Either of these two accounts (that differences in thought give rise to differences in language, or that a third variable gives rise to both) may explain why some research has suggested that even when languages differ in the meanings of words, speakers’ underlying concepts are largely the same. This view has been supported by findings that in some cases differences in language do not seem to correspond with differences in thought. For example, even though English, Spanish, and Chinese differ in how they attribute word labels to storage containers, in prior work speakers of these three languages have made comparable similarity judgments about containers [82]. Of course, it is also possible that differences in language can cause differences in thought, but that linguistic differences dont always do so, but existing cross-linguistic research is not poised to address this possibility.

Some support for the account that language can shape thought comes from research in which language is manipulated within a single cultural group. Such work has shown that behavior can differ systematically with linguistic manipulations. This approach rules out the possibility that cross-linguistic language and thought findings result solely from pre-existing cross-cultural differences in thought and also rules out the possibility of a third variable that shapes language and thought, since it eliminates cultural differences between language conditions (by randomly assigning participants to
A language manipulation within a single cultural group has shown, for example, that language can facilitate category learning. In one study, participants learned to distinguish approachable and non-approachable creatures, which could be learned on a visual basis alone [80]. However, when participants also learned linguistic labels for the creatures, they learned the categories more quickly.

Similar approaches have shown that relational language in particular seems to help people discover abstract commonalities, leading to clearer mental representations [45]. In one study, children watched as a star was placed behind a card either on the top, middle, or bottom shelf [77]. They were then asked to find the star in a different shelving unit, almost identical to the first. Performance was improved when the researchers used relational words like on, under, top, or bottom while hiding the star, suggesting that the relational words actually helped children create a more precise representation of where the object was in memory.

Additional research has shown that teaching children specific linguistic constructions improves theory of mind understanding [60]. In one experiment, some children were trained on sentential complements, which allow tensed propositions to be embedded under a main verb. Those propositions may have independent truth value, such as in the sentence John said that Mary went shopping, where the main clause can be true (John said X) while the embedded clause is false (Mary went shopping). Other children were trained on relative clauses, which also embed clauses, but do so under a noun (The boy that had red hair...), whereas sentential complements do so under a verb. Children who were trained on sentential complements showed an increase in their theory of mind understanding, as demonstrated by performance on a false belief task, that was greater than the change in theory of mind understanding for children trained on relative clauses. This result suggests that acquiring a linguistic construction (specifically, one that uniquely
allows for the representation of false beliefs) can open up new ways of understanding the world.

These studies show that language can improve abilities to categorize objects, remember spatial relations, and understand that others’ beliefs may be different from our own. However, they have a common theme which also differentiates them from most cross-linguistic work: these studies show that the introduction of a term or construction improves performance on certain tasks compared to when the term or construction is absent. Cross-linguistic work, on the other hand, doesn’t solely focus on presence or absence of a linguistic feature, but instead also investigates behavior in speakers of languages that do have a way of encoding some feature of the world, but have differing ways of doing so. For this reason, existing research cannot directly address the question of whether patterns in language can cause differences in the way speakers of different language represent the world.

1.2.2 The Nature of Effects of Language on Thought

Even if language can cause people to perform differently on related tasks, what is the nature of that effect? To what extent can something learned through language be nonlinguistic in mind?

One side of this debate includes claims that effects of language on thought are shallow and remain on the surface, only influencing cognition that is decidedly linguistic. For example, some scholars have interpreted Dan Slobin’s [107] thinking for speaking account as consistent with this claim. Thinking for speaking posits that when we prepare to use language (whether productively, for example by speaking or writing, or receptively, for example by listening or reading), we encode the world in a manner that our language requires in order to verbalize it. It is possible that effects of language on behavior arise only when the task explicitly requires language involvement [53].
Research in the domain of motion events supports the possibility that language shapes behavior specifically when language is also implicated in the task used to measure behavior. For example, Papafragou, Hulbert, & Truesdale [97] measured Greek and English speakers’ eye movement patterns as they watched motion events. When the participants were instructed to watch them in preparation for describing them, eye movements reflected linguistic patterns: Greek speakers were much more likely to focus on the path of the movement (Greek verbs predominantly specify paths of motion) than English speakers (English verbs are more likely to specify the manner of motion). When the participants were not instructed to watch the events in preparation for explaining them, however, these cross-linguistic differences in eye movements disappeared.

Another approach that has been used to understand whether effects of language on thought are necessarily linguistic has been to apply a verbal interference task. Interference tasks rest on the assumption that a cognitive process that requires people to engage their phonological loop will be disrupted by the addition of another task that recruits verbal working memory resources in the moment. Common verbal interference tasks require people to remember linguistic information, like a series of digits or letters, while also completing the main task of interest.

The application of verbal interference has suggested that the effect of language on color discrimination is online [127]. In other words, the interaction between lower-level perceptual processes and higher-level linguistic processes leads to the color processing advantage for Russian speakers across the siniy-goluboy boundary. Follow-up work has further supported the conclusion that effects of language on color perception arise through the online recruitment of linguistic cognitive resources. For example, cross-boundary advantages for color distinctions only occur when the colors are presented to the right visual field (and are therefore projected to the left hemisphere, the locus of most linguistic processes in most people [52]; for review, see [101].
Thus, a number of findings have led some researchers to suggest that when language shapes behavior, the effects are of language on language [53]. These researchers see the role of language as one with superficial, unnecessary, and inconsistent effects on thought [98]. Proponents of this account dismiss the possibility that language may exert effects on nonlinguistic thought [90].

However, on the alternative end of the theoretical spectrum, other researchers assert that effects of language on thought may be deep and persistent, extending beyond the linguistic realm. One type of work supporting this claim includes studies that use entirely nonlinguistic methods when observing differences in thought that are consistent with differences in languages. For example, using a task that measures reaction times for making judgments about sequences of images by pressing buttons, researchers have observed that speakers of different languages perform in ways that are consistent with patterns their languages [13]. Specifically, Chinese speakers show a stronger top-to-bottom bias for thinking about time, consistent with their natural language metaphors, that is greater than English speakers show.

These studies provide a glimpse into the language-and-thought corner of Cognitive Science research, with a foundation of cross-linguistic studies showing systematic correspondences between patterns in natural language and mental representations. However, two theoretical questions remain open: Can language really cause us to see the world in new ways? And if so, to what extent can language create nonlinguistic knowledge?

1.3 Spatio-Temporal Metaphors: Bridging Case Study

Of all the conceptual domains we might understand metaphorically, the best studied is time. Within a given culture or community, individuals think and talk about time in ways that are both stable and shared. This often involves using space to structure
both temporal speech and temporal understanding (for reviews, see [12, 93]).

In English, the way people talk about time strongly overlaps with the way they talk about space, with many of the same words and constructions used to talk about the relationship between objects in space and events in time [28, 74]. Just as we might say that an bird flew by, we can say that an hour flew by. We can push forward a truck or a meeting, believe that a fire or a winter is behind us, or worry that a bear or a deadline is approaching. Prior work has demonstrated that people don’t just use spatial terms to talk about time; they also appear to use specific spatial representations when thinking about time.

For example, in one early set of studies, participants answered questions about time differently, depending on how they had just been thinking about space [9]. Some participants were asked to imagine themselves moving through space toward a goal, while others imagined an object coming toward them instead. Then they were asked a seemingly unrelated question about time. They were told that Next Wednesdays meeting has been moved forward two days. What day is the meeting now that it has been rescheduled? The phrase moved forward is ambiguous in this context [88, 84]. If you imagine yourself moving forward in time, then forward is further in your direction of motion (from Wednesday to Friday). But if you imagine time moving toward you, then forward can be further in the direction that time is moving (from Wednesday to Monday). Participants who had earlier imagined themselves moving forward through space were more likely to say the meeting would be on Friday (further in their direction of motion) than those who had imagined an object traveling toward them. These results suggest that when interpreting spatio-temporal metaphors, people draw on the analogous spatial representations they have readily available in mind.

Further, people appear to construct metaphor-consistent spatial representations for time even when not processing spatio-temporal metaphors in language. For example,
English metaphors commonly place events on a sagittal mental timeline with the future ahead or in front and the past behind. Correspondingly, English speakers are faster to make decisions about events when doing so involves moving their arm or pushing a slider forward to indicate that an event is in the future and pulling their arm or slider back towards the body to indicate that an event is in the past than for the reverse mapping [104, 121]. They are also faster to make past-future judgments to words shown in front of an image of a person for future concepts and behind for past concepts [117]. In addition, people are faster to step forward when they hear words related to the future than the past and faster to step backward for the past than the future [102]. In a deliberate gesture task, they gesture to the front when talking about the future and the back when talking about the past [20].

These studies suggest that English speakers have representations of time that align with spatio-temporal metaphors in English. As discussed previously, however, this does not necessarily mean that metaphors in language played any role in constructing these representations. Patterns in language may simply reflect representations that have emerged through some combination of physical experience and innate predisposition. To examine this possibility, researchers have compared representations of time in speakers of different languages, focusing on cases where different languages rely on different spatial metaphors for time.

For example, English uses primarily front-back metaphors that place the future in front of us (i.e., looking forward to the future) and the past behind (putting the past behind us). In Aymara, this pattern is reversed, so that past events are in front and future events are behind [94]. Consistent with their linguistic metaphors, Aymara speakers gesture about future events as behind them and past events as in front (in contrast to Spanish speakers from the same region whose gesture patterns and metaphors place the future in front, as in English) [94].
Researchers have also empirically examined cross-linguistic differences in the representation of time on the horizontal and vertical axes. While English relies primarily on horizontal (front/back) spatial terms for talking about time, Mandarin commonly uses both horizontal and vertical (up/down) terms. In Mandarin, earlier events are said to be *up* and later events are *down*. Although English does occasionally use vertical metaphors (like *passing knowledge down the generations*), these metaphors are not as systematic, productive, or frequent as they are in Mandarin [103]. This difference between Mandarin and English in their use of the vertical axis to talk about time has provided opportunities to investigate corresponding differences in thinking about time.

A variety of experimental methods have revealed that Mandarin speakers are more likely to create vertical representations for time than English speakers are, both in linguistic and non-linguistic tasks (e.g., [10, 13, 44, 87, 59, 73, 131]). For example, Mandarin speakers are more likely than English speakers to arrange picture sequences vertically when asked to indicate time progression [87], more likely to place events on the vertical axis in an elicited pointing task [11, 44, 73], and more likely to make spontaneous co-speech gestures on the vertical axis when talking about time [59]. The tendency to arrange time vertically is modulated by a number of linguistic variables: it is stronger when participants are tested in Mandarin, in participants who are more fluent in Mandarin, and when the task includes vertical metaphors in Mandarin [44].

Non-linguistic tasks that use pictures as stimuli and button presses as responses have also revealed that Mandarin speakers have an active earlier-up implicit association between space and time [87, 13, 44]. In the task used by Boroditsky et al. [13] and Fuhrman et al. [44], which was adapted for this current work, participants see an initial image (e.g., Julia Roberts in her 20s) followed by a second image (e.g., either a younger or older Julia). Their task is to determine whether the second image shows a conceptually earlier or later time-point than the first and press a key corresponding to their decision.
The response keys are arranged so that the earlier key is either above or below the later key. On these tasks, Mandarin speakers show an implicit association between earlier time points and higher space, responding faster when the earlier key is above the later key than the reverse. English speakers tested on the same task show either no evidence of a specific space-time association on the vertical axis or a significantly weaker effect than Mandarin speakers.

The findings discussed so far establish differences in temporal representations between speakers of different languages. Are linguistic metaphors in part responsible for these differences in thought? Of course, metaphors in language are not the only source of space-time mappings. For example, in addition to representations of time on the front-back axis that align with linguistic metaphors, English speakers also show robust representations of time on the left-right axis, with earlier or past events on the left and later or future events on the right. There are no widely used left-right metaphors in English; we do not say things like Tuesday is left of Wednesday. Instead, these left-right patterns appear to emerge from visuo-motor cultural practices in reading and writing. Speakers of Hebrew and Arabic, who read and write from right to left, show right-to-left mental timelines (e.g., [120, 43, 96]; see also [8]). These findings suggest that cross-cultural differences in thinking about time can come about through sources other than spatio-temporal metaphor, including practices like reading and writing.

### 1.3.1 Does Language Play a Causal Role?

The cross-linguistic work reviewed so far shows that speakers of different languages represent time differently, in ways that appear to correspond to the patterns of space-time metaphors in their language, and covary with aspects of their linguistic experience (e.g., how fluent they are in the language of interest). As discussed earlier, simply establishing such correlations does not establish that linguistic metaphors play a
causal role in constructing different representations of time. It is possible, for example, that both the differences in representations and differences in metaphors arise because of some other set of underlying cultural differences, unrelated to language. Whenever studies compare behavior across different linguistic groups, the design is necessarily quasi-experimental. It is not possible to randomly assign participants to be English or Mandarin speakers - they are already English or Mandarin speakers when they participate, and with these differences come a potentially infinite set of other confounding factors.

How then, can we establish whether patterns in metaphorical language can indeed shape how people think?

A few researchers have looked to metaphor training studies as an opportunity to investigate whether metaphors can cause people to think in consistent ways by randomly assigning participants to groups that learn new metaphors. For example, in one study, participants read about either the past or the present as metaphorically heavy (The past carries particular weight for who you are today. You must carry your past with you wherever you go. or The decisions of your past carry no weight. It is your decisions today that define who you are, and you must hold the present with great care. [106]). In an ostensibly separate study, they estimated a book’s physical weight. Those who had read that the past is heavy judged older books as physically heavier (than newer books), while those who read that the present is heavy judged newer books to be heavier (than older ones). The authors concluded that experience with a novel and newly-learned metaphor shaped the way people mentally associated abstract properties with concrete objects.

In other research, participants have learned new spatial metaphors to talk about musical pitch [36]. The training work was inspired by cross-linguistic differences in talking about pitch: while Dutch speakers describe pitches as either high or low, Farsi speakers describe them as either thick or thin. Dutch and Farsi speakers’ performance on two pitch-reproduction tasks that used nonlinguistic stimuli and responses reflected their
different metaphor patterns in language. Then, Dutch speakers learned to describe pitch in terms of thickness, and their performance on the same tasks mirrored that of Farsi speakers in the cross-linguistic work. Thus, this work not only found that people who use different pitch metaphors think about pitch differently, but also that, at least in the domain of pitch, linguistic metaphor can play a causal role in shaping the way people think.

The studies in this dissertation are also centered around a metaphor training paradigm, as a way of investigating whether linguistic metaphors for time can create new mental representations, and whether these representations may be nonlinguistic in nature.

1.3.2 Mechanisms: Testing Verbal & Spatial Working Memory with Dual Tasks

One common way of assessing whether specific working memory resources are involved in some other cognitive process is to use a dual task. Dual tasks leverage the limited capacity of human working memory to investigate whether the same cognitive resources are required for processing distinct streams of information (e.g., [5, 99]. Dual tasks commonly tax either spatial or verbal working memory.

In a common verbal interference paradigm, participants encode a series of letters or digits and maintain them in memory while doing a main task of interest. Performance on either the verbal task, main task, or both will be altered if the two tasks require the same working memory resources to be completed. Prior work using verbal interference as a probe has uncovered striking examples of involvement of linguistic routines or representations in a variety of cognitive tasks that we previously had not expected to involve language. For example, verbal interference specifically changes people’s performance (in ways predictable from patterns in language) when making basic color judgments, counting dots, performing simple algebra, reorienting in a small room, and
judging the similarity of motion events [127, 42, 41, 66, 3]. Taken together these findings suggest that even when people are not required to use language to speak or understand, normal human cognition is in fact language-augmented cognition. Human brains recruit linguistic processes in a wide variety of seemingly non-linguistic tasks, in a way that has real consequences for cognition and behavior.

The Corsi working memory task [30] is often used to assess whether visuospatial working memory is necessarily recruited when people complete a different task. In the traditional computerized Corsi working memory task, squares on a screen light up one at a time, and a participant must remember which squares lit up and in which order. Prior work has demonstrated that completing a Corsi trial requires active rehearsal of locations by shifting spatial attention [132].

The Corsi task is often paired with other tasks. That is, participants see boxes light up (they encode the stimuli), maintain that spatial information in memory while doing another task, and then they finally report which boxes lit up at the start of the trial. Prior research has found that either Corsi performance or performance on the intervening task is impaired when the Corsi task is paired with a visual search task [95], a tapping task during Corsi encoding, hitting a button at random intervals [123], touching targets on a physical board [108, 35], physically moving [100], reading visually-presented words, or hearing tones and reporting where they came from [108]. The tasks that interfere with the cognitive processes needed to complete the Corsi task are varied, but they all require spatial working memory [108].

If either spatial or verbal working memory is required in the moment for associations to manifest, then taxing the relevant working memory system should affect those associations. A dual task may disrupt participants’ ability to conjure up canonical associations at all, such that they no longer demonstrate those associations in behavior. Using congruency effects as the indicator of an association, a dual task would, in this
case, reduce or eliminate congruency effects. Throughout this dissertation, this possibility is referred to as the *reduction* effect.

However, it is also possible that a working memory task might interfere with performance on a main task by making incongruent trials especially difficult, as participants must maintain the interfering information, their active association, and the incongruent task instructions in mind all at the same time. Throughout, this possibility is referred to as the *amplification* effect.

Of course, a final possibility is that working memory resources are not necessarily recruited in the moment when associations are observed in behavior. Instead, such associations are, at least to some extent, stored in long-term memory, so that they manifest even when spatial or verbal working memory resources are already tied up. This possibility is the *long-term memory* account.

### 1.4 Overview of Studies

The following studies use the domain of spatio-temporal metaphors to investigate overarching theoretical questions at the intersection of two major lines of cognitive science research:

**Metaphor**: How do we build abstract knowledge? Do the analogical relations implied in systematic natural language metaphors guide how we construct mental representations? **Language & Thought**: Can patterns in language actually cause us to think in new ways? And can things learned through language be represented non-linguistically?

Experiment 1 (Ch. 2) takes these questions up by specifically asking: Does learning a new system of linguistic metaphors for talking about time change the way people think about it? To this end, participants learned vertical spatio-temporal metaphors (i.e., *breakfast is above dinner* or *breakfast is below dinner*) before completing a non-
linguistic space-time association task. That work also investigated the nature of the newly-learned representation – is it linguistic? – by applying verbal interference as participants completed the main task.

In the first experiment, participants showed mental representations consistent with their newly learned metaphors. That is, people who learned that earlier events are above later ones also showed an earlier-is-up implicit mental timeline, while those who learned the opposite metaphors showed the opposite. Further, these associations were unaffected by verbal interference, suggesting that the newly learned representations were nonlinguistic in nature, despite their linguistic origin in the experiment. The following studies build on these preliminary findings.

Experiment 2 investigates one alternate explanation for the previous finding that new space-time associations were robust under verbal interference: what if the verbal dual task we used doesn’t actually interfere with linguistic representations? Thus, we replicated prior work [42] showing that verbal interference disrupts American college students’ abilities to count objects, while spatial interference does not. By ensuring that the verbal dual task does interfere with a known linguistic routine, we have more support for the conclusion drawn from Exp. 1 – that newly learned metaphors can shape thought by creating nonlinguistic representations.

The next chapter (Ch. 3) compares the representations participants acquired through short-term linguistic training in the lab to those that are consistent with natural language metaphors; namely, in Mandarin Chinese speakers. In Experiment 3, Mandarin-English bilinguals completed the same non-linguistic time judgment task as in Exp. 1 with the same interference conditions, but did not learn a new system of metaphors. We compared their performance to English speakers who completed the same task, and to the English speakers in the first experiment, who learned new systems of metaphors.

Next, we investigated simultaneous long- and short-term language influences on
mental representations of time. In Exp. 4, Mandarin-English speakers learned the new vertical metaphors in the lab (in English, as in Exp. 1) and completed the same tasks as in the prior experiments. Together, the two experiments in Ch. 3 allow us a clearer understanding of how representations arising from newly learned metaphors in the lab interact with and compare to those that have arisen from a lifetime of natural language use.

The subsequent chapter (Ch. 4) explores potential differences in nature between mental representations of time that arise from nonlinguistic cultural experiences, those on the lateral axis, and those originating in language. Experiment 5 included the same stimuli and procedures as the previous experiments, but responses were oriented laterally, instead of vertically. Experiments 6 and 7 build on the prior experiment by investigating lateral mental number lines in the same paradigm. While the former (Exp. 6) required participants to make two time or number judgments while maintaining interference information in mind, the latter (Exp. 7) required them to make 16 judgments, to test whether dual tasks may be more likely to interfere when the load must be remembered for a longer time and over more subsequent judgments.

The final experiment using these same materials (Exp 8) addresses one possibility for the lack of spatial interference in all of the prior experiments: perhaps general spatial working memory resources are not recruited when people invoke a mental time, but instead specific, axis-relevant spatial working memory resources are. To this end, participants made temporal judgments, again with lateral responses, while undergoing interference concentrated either on the vertical or the horizontal axis.

Experiments 1-8 are then included in a meta-analysis to allow us to explore commonalities and differences across the datasets (Ch. 5). For example, what is the time-course of mental time- and number lines? In other words, how quickly do participants adapt to incongruent mappings, and does this adaptation process differ based on axis,
domain, or representation source?

Chapter 6 moves into a partially new experimental paradigm, to address whether representations newly learned from language can manifest in spontaneous behavior, specifically through co-speech gestures. In Experiment 9, participants learned the same new vertical metaphors for time used in prior experiment, and then described temporal concepts. We analyzed their gesture patterns to see whether the new vertical representations integrated with pre-existing mental timelines in a way that would seep out in gestures.

The final experimental chapter (Ch. 7) investigates the reverse relationship: can implicit mental representations of time become encoded in linguistic metaphors? Experiments 10 and 11 document a pattern of language use present in a sub-community of American military speakers; namely, the use of left-right metaphors (like move the meeting right two days to mean move the meeting later.). Although these linguistic metaphors were not unintelligible to civilian participants, they were notably less acceptable to this group, suggesting that the patterns we observed in military members indeed reflect a shift from previously nonlinguistic mental associations to linguistic conventions.

Chapter 8 includes a concluding discussion to summarize and synthesize the results of the experiments included in this dissertation.
Chapter 2

New Space-Time Metaphors Foster New Non-Linguistic Representations

2.1 Abstract

What is the role of language in constructing knowledge? This chapter explores whether learning new relational language can create new ways of thinking, and if so, whether those new patterns result from the use of online linguistic routines. To this end, we taught English speakers to talk about time using new vertical linguistic metaphors, saying things like breakfast is above dinner or breakfast is below dinner, and then asked whether the newly-learned representations persist under verbal interference. We found that learning new metaphors created new space-time associations that could be detected in a non-linguistic implicit association task, and these newly learned representations persisted under verbal interference. This chapter shows that learning new relational language can be a powerful tool in constructing new representations and expanding our cognitive repertoire.¹

2.2 Experiment 1 Introduction

Do the analogical relations implied in natural language metaphors guide how we construct our knowledge? Much previous work has singled out analogical learning as a key mechanism in the human ability to construct complex knowledge (e.g., [46, 68]), and has demonstrated the role of relational language in particular in conceptual development (e.g., [47, 25]). Here we ask, does the structure of our mental representations depend in part on the analogical relations suggested by the metaphorical systems in our languages?

We take up this question in the example domain of time, specifically people’s representations of the temporal order of events. We ask whether learning new linguistic metaphors for talking about time creates new cognitive patterns that can be detected in nonlinguistic tasks. Further, what is the nature of these newly learned representations? Do people create new linguistic routines that "run alongside" other representations? Or can learning relational language create new non-linguistic patterns in thought?

2.2.1 Spatio-Temporal Metaphors: A Brief Overview

As discussed in the Introduction (Ch. 1), the ways people talk about time strongly overlap with the ways they talk about space [28, 74]. Extensive research has shown that people don’t just talk about time using spatial words; they also appear to use specific spatial representations when thinking about time.

For example, in one early set of studies, participants answered questions about time differently, depending on how they had just been thinking about space [9]. If they had imagined themselves moving through space to a goal, they were likely to answer an ambiguous temporal question (Next Wednesdays meeting has been moved forward two days. What day is it now?) with a consistent ego-moving response (Friday). If they had imagined an object toward them, they were more likely to respond consistently with a
time-moving schema (Monday).

Other experiments have shown that even when not processing spatial language, English speakers think in ways consistent with English temporal metaphors (referring to the past as *behind* and future as *ahead*; [117, 104, 121, 102, 20]). These studies suggest that English speakers have representations of time that align with spatio-temporal metaphors in English.

Cross-linguistic research has demonstrated that these findings for English speakers are not universal; instead, speakers of other languages also hold mental representations of time that are consistent with their own language patterns. For example, Aymara speakers, whose metaphors reverse English conventions (in Aymara, the past is in front and the future behind) also seem to think in ways consistent with their metaphors, gesturing about the past in front and future behind them [94].

Mandarin Chinese has provided another opportunity for cross-linguistic comparisons of mental representations of time. In Mandarin, earlier events are said to be *up* and later events are *down*. A variety of experimental methods have revealed that Mandarin speakers are more likely to create vertical representations for time than are English speakers, both in linguistic and non-linguistic tasks (e.g., [10, 13, 44, 87, 58, 73, 131]).

One non-linguistic task that uses pictures as stimuli and button presses as responses, used by Boroditsky et al [13] and Fuhrman et al [44] was adapted for this current work. In that task, participants see an initial image (e.g., Julia Roberts in her 20s) followed by a second image (e.g., either a younger or older Julia). Their goal is to determine whether the second image shows a conceptually earlier or later time-point than the first and press a key corresponding to their decision. The response keys are arranged so that the earlier key is either above or below the later key. On these tasks, Mandarin speakers show an implicit association of earlier=up, responding faster when the earlier key is above the later key than the reverse. English speakers tested on the same task
show either no evidence of a specific space-time association on the vertical axis or a significantly weaker effect than Mandarin speakers.

The cross-linguistic work reviewed so far shows that speakers of different languages represent time differently, in ways that correspond to the patterns of space-time metaphors in their language. Of course, simply establishing such correlations does not establish that linguistic metaphors play a **causal** role in constructing different representations of time. Whenever studies compare behavior across different linguistic groups, the design is necessarily quasi-experimental. It is not possible to randomly assign participants to be English or Mandarin speakers - they are already English or Mandarin speakers when they come in, and with these differences come a potentially infinite set of other confounding factors. How then, can we establish whether patterns in metaphorical language can indeed shape how people think?

### 2.2.2 Experiment 1 Overview

In this work, we experimentally created two new “language communities.” We taught English speakers new metaphors for time, with one group learning to say things like *Tuesday is above Wednesday* and the other group learning to say things like *Tuesday is below Wednesday*. We examine whether learning new metaphors can create the same kinds of differences in implicit associations between space and time as have been found across cultures. Thus, participants completed a nonlinguistic time judgment task that has been used in prior work to establish cross-cultural differences in representations of time [43, 44].

Further, we investigate the locus of these effects. Do the newly learned mappings reside in the linguistic sphere? Or are they nonlinguistic representations (ones that are not disrupted by verbal interference)? To this end, some participants completed the task without any secondary task, while others did it under conditions of either verbal
interference or spatial interference (as a control).

Prior work using verbal interference as a probe has uncovered striking examples of linguistic routines or representations being involved in a variety of cognitive tasks that we previously did not expect to involve language. For example, verbal interference has been found to specifically change people’s performance (in ways predictable from patterns in language) when making basic color judgments, counting dots, performing simple algebra, reorienting in a small room, and judging the similarity of motion events [127, 42, 41, 66, 3]. Taken together these findings suggest that even when people are not required to use language to speak or understand, normal human cognition is in fact language-augmented cognition. Human brains recruit linguistic processes in a wide variety of seemingly non-linguistic tasks, in a way that has real consequences for cognition and behavior.

Some scholars have suggested that this in the moment meddling is the only way in which experience with language can influence cognition (e.g., [90, 98]. That is, we can create new linguistic representations and processes that run alongside non-linguistic representations, but experience with language does not create new non-linguistic representations or processes. Here we ask whether this is indeed the case. Can learning a new way of talking act as a formative force, creating new non-linguistic representations that will continue to manifest in cognition and behavior even when our ability to recruit linguistic resources is disrupted?

2.3 Experiment 1 Methods

2.3.1 Participants

A total of 206 UC San Diego undergraduates were recruited through the psychology recruitment system and participated for payment or course credit.
2.3.2 Procedure

In this experiment, all participants learned a new way to talk about time (see Metaphor Training for details). After training, participants completed a time judgment task (detailed in Time Judgments section). We used the same non-linguistic space-time implicit association task that has been used to establish cross-cultural differences in time representations between English, Mandarin, and Hebrew speakers in prior work [13, 44, 43]. In this task, participants see sequences of two photos and indicate whether the second image shows a conceptually earlier or later time point than the first. Responses are made using button presses, and in different conditions the earlier key is located either above or below the later key. Participants completed these judgments under conditions of either verbal, spatial, or no interference. For the spatial and verbal interference conditions, participants first completed a calibration block (see Interference Calibration for details) to ensure that the interference difficulty was appropriately scaled across individuals.

The study proceeded in the following phases: (a) interference calibration block (shown in Figure 2.1), (b) metaphor training block, (c) time judgment block (shown in Figure 2.2), (d) second metaphor training block (identical to the first, used as a refresher), and (e) a final time judgment block. For each participant, the first and second time judgment blocks had opposite assignments of response keys; in one block, the earlier key was above later and in the other the earlier key was below later. This made it so that each participant completed one time judgment block for which the arrangement of the response keys was congruent with their metaphor training and one for which it was incongruent. Which metaphor system people were trained on and whether the first time judgment block was congruent or incongruent with the training were counterbalanced across participants.
**Interference Calibration**

Participants in the verbal and spatial interference groups first completed a calibration block to ensure that the interference tasks were properly tuned for individual ability. The materials and calibration procedure were identical to those used in Frank et al [42]. A full description of the procedure is available in Frank et al [42]. We kept the procedures identical to this prior work because it provides a clear precedent, showing that this same verbal interference task, calibrated in this same way, strongly interferes with people’s ability to engage a linguistic routine like counting. If our congruency effects are also due to participants employing an online linguistic routine, then we should find the same deleterious effects of verbal interference.

In the verbal interference condition, participants saw a string of letters, which they held in memory as they completed a visual search task. After they responded to the visual search trial, they typed the letters they saw initially. In the spatial interference condition, they saw a 4x4 grid of white squares, and a number of boxes turned blue one at a time. They did the same visual search task, and finally received a blank 4x4 grid and clicked on the boxes that had turned blue. Each participant started with a load of 2 items (either two letters or two boxes), and the difficulty was calibrated in a 2-up, 1-down staircase. If participants got two trials correct at a given difficulty level, the number of items was increased by one for the next trial, and if they got a trial incorrect, the number of items was decreased by one for the next trial. Calibration continued for 60 trials, and the final number of distractor items a person experienced was used throughout the main task.

**Metaphor Training**

After interference calibration (or at the beginning of the experiment, for participants in the no interference condition), participants were told that they would learn a
Figure 2.1: Interference Calibration. Participants began by seeing two boxes turn blue or two consonants, one at a time. They then searched for an L among an array of Ts, pressed the L key if the L was present (K otherwise), and recalled the interference array by either clicking on boxes or typing in consonants.

new way of talking about time. First they read five example sentences showcasing this new system of talking. For half of the participants, earlier events were said to be above or higher and later events were said to be below or lower (e.g., Thursday is higher than Friday; When you eat breakfast, dinner is below you.). This system was reversed for the other half of participants so that earlier events were below/lower and later events were above/higher (Thursday is lower than Friday; When you eat breakfast, dinner is above you.).

Participants then completed 90 training trials to practice this new system. On each trial, they saw a new sentence describing a temporal relation (e.g, Lincoln was president ________ than Carter) and were asked to type in the correct spatial term (selecting from the two options provided: higher and lower for the previous example). If the answer was incorrect, the computer provided feedback and required the participant to correct their answer. The keyboard was flat on the surface of the table so that participants could type normally.
Time Judgments

We used the same space-time implicit association task as has been used in prior work [13, 44, 43]. Participants pressed a central key to begin the trial, and the first of two images appeared on the screen (e.g., Julia Roberts in her 20s). After 2 seconds, this image was replaced by the second image (e.g., either a younger or older Julia). Their task was to determine whether the second image showed a conceptually earlier or later picture than the first and press a key corresponding to their decision. For all participants, the earlier key was above the starting key in one block, and the later key was above the starting key in the other. This meant that in one block the key mapping of responses was spatially congruent with the system of metaphors participants had learned, and in one block the key mapping was incongruent. In this portion of the experiment, participants made responses on a keyboard mounted vertically (perpendicular to the table surface). All participants responded with their right hand.

At the beginning of each time judgment block, participants completed 10 practice trials with feedback. After the practice trials, there were 56 experimental trials without feedback. For both practice and experimental trials, image sequences were selected in a random order for each participant. Every sequence was shown only once in each block, and whether the earlier or later picture was shown each time was random (i.e., if the earlier picture in the Julia Roberts sequence was shown in the first block, the later picture in the sequence was not necessarily shown in the second block). After completing the time judgments, participants returned to another block of training on the same metaphor, and then did another time judgment block (this time with the key mapping reversed from their first run). At the end of the experiment, we collected information about participants’ language backgrounds.
Figure 2.2: Task with Interference. Participants in the spatial and verbal interference conditions saw boxes turn blue or a string of consonants flash, one at a time. They then completed a time judgment trial. Finally, they recalled the interference distractors they had seen at the beginning of the block. Participants in the condition without interference completed only the time judgment trials.

2.3.3 Participant & Trial Inclusion Criteria

Prior work including similar methods has used between 40 and 118 participants per condition [43, 13, 44]. Following this work, we included 64 participants for each interference group, including 32 who learned each of the two metaphor systems. This allowed for a fully counterbalanced sample across interference types, metaphors learned, and block order.

There were 11 participants with accuracies below 25% for one time judgment block. We assumed that they reversed the instructions for that block despite completing ten practice trials with feedback, so these participants were excluded. We also excluded participants whose accuracies on one or more blocks of the training or time judgment task were lower than 3 standard deviations from the sample mean accuracy. This resulted in eliminating 13 participants for low metaphor training accuracies and 8 additional participants for low time judgment accuracies. The mean accuracies for verbal and
spatial interference were determined separately, and participants who performed below 3 standard deviations from the sample mean accuracy for their interference type were eliminated. This resulted in eliminating 2 participants from the verbal interference condition and one from the spatial interference condition.

Only trials for which the time judgment was correct (94.0% of all trials) and interference memory was correct (72.1% of all interference trials) were included in analyses. In addition, only reaction times within 3 standard deviations of each participant’s cell mean (98.0% of correct trials) were included.

Participants who were excluded were replaced to result in a total of 192 participants for a fully counterbalanced design.

2.3.4 Analyses

We fit the data with mixed effects models using Laplace Approximation using the lmer() function within the lme4 analysis package in R [7]. The full mixed effects model included time judgment, training, interference conditions, and block order (whether the congruent block was the first or second time judgment block) as fully crossed fixed effects, and subjects and items as random effects.

To examine main effects and interactions, we constructed reduced models that were identical to the full model but did not include the main effects or interactions of interest and compared these reduced models to the full model. Both models had random intercepts plus random slopes for the within-subjects variable of task condition.
2.4 Experiment 1 Results

2.4.1 Main (Hypothesis-Relevant) Results

Learning new spatio-temporal metaphors created new non-linguistic representations of time. That is, English speakers adopted vertical representations for thinking about time that were consistent with the system of metaphors they learned in the lab. These metaphor-consistent representations were not disrupted by either verbal or spatial interference.

Participants were faster when the time judgment response keys were arranged congruently with their newly learned metaphors (1559 ms) than when the responses were arranged incongruently (1619 ms), confirmed by a significant interaction between training and time judgment conditions, $X^2(1)=8.47$, $p=.004$. There were no speed-accuracy tradeoffs, as participants in the congruent conditions were faster and numerically slightly more accurate. Table 2.1 shows mean reaction times for each group of participants in each time judgment block.

The congruency effects did not differ between the no interference, verbal interference, and spatial interference conditions; there was not a 3-way interaction between interference type, training condition, and time judgment condition, $(X^2(2)=1.31$, $p=.52)$. Pairwise comparisons of the congruency effect between all possible interference types revealed the same null effect, (between verbal and no interference, $X^2(1)=1.44$, $p=.23$; between spatial and no interference, $2(1)=0.79$, $p=.37$; between verbal and spatial interference, $X^2(1)=0.02$, $p=.88$. Figure 2.3 shows results across the interference conditions.

There was no difference in overall reaction times between the groups that learned the two metaphor systems $(X^2(1)=0.007$, $p=.94)$. 

34
**Table 2.1:** Metaphor Training Reaction Times by Interference Type (SD). All reaction times are in ms.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learned earlier is up</td>
<td>1500 (414)</td>
<td>1662 (390)</td>
<td>1472 (391)</td>
</tr>
<tr>
<td>earlier key above</td>
<td>1517 (396)</td>
<td>1769 (412)</td>
<td>1604 (370)</td>
</tr>
<tr>
<td>Learned earlier is down</td>
<td>1432 (338)</td>
<td>1759 (333)</td>
<td>1635 (458)</td>
</tr>
<tr>
<td>earlier key above</td>
<td>1392 (333)</td>
<td>1698 (301)</td>
<td>1630 (396)</td>
</tr>
</tbody>
</table>

People were equally accurate on congruent (94.0%) and incongruent (93.9%) trials, suggesting that results in response times were not due to speed-accuracy trade-offs; including congruency in the analysis as a factor (instead of inferring congruency from the interaction between training and task conditions), $F(1,186)=0.07$, $p=.80$.

There was a significant negative correlation between trial number and the congruency effect ($r(54) = -0.32, p = .02$). This finding suggests that the instructions for a block of time judgments act as a separate kind of training, strengthening the instruction-consistent representation of time over the course of a block.

### 2.4.2 Ancillary Results

#### Reaction Times

There was a main effect of interference type on overall reaction times during the task ($X^2(2)=18.4$, $p = .0001$). Pairwise comparisons showed that participants without interference were fastest ($M = 1460$ ms). They were faster than those in the spatial interference condition ($M = 1585$ ms; $X^2(1)= 4.46$, $p = .04$), who were in turn faster than those in the verbal interference condition ($M = 1722$ ms; $X^2(1) = 4.88$, $p = .03$).

Congruency effects were not moderated by the number of interference distractors people experienced ($r = .13, p = .15$).
Figure 2.3: Results of Experiment 1: Congruency effects were calculated for each participant by subtracting their mean reaction time on trials congruent with their newly learned metaphor from their mean reaction time on trials incongruent with their newly learned metaphor. Error bars show SEMs.

Accuracies

During the metaphor training, participants who learned that earlier events are below later ones were 2% more accurate (95.9%) than those who learned that earlier events are above later ones (93.9%); t(181)=5.02, p = .000001. This small difference may reflect a common intuition among English speakers that the future should be above rather than below [61]. It should be noted that this intuition is the opposite from the system of metaphors in Mandarin (which places the past above).

Time judgment accuracies are presented in Table 2.2. Aside from a marginal difference in the time judgment accuracy across interference types (those without interference were more accurate than those with, F(2,186) = 2.77, p = .07), there were no
other differences in task accuracy across conditions: for main effects of training and task conditions, and all interactions among training condition, task condition, and interference type, all ps > .11).

Because the number of interference distractors was calibrated for each participant, those in the spatial interference group completed the time judgments with an average of 4.09 (SD = 1.3) boxes, and those in the verbal interference group saw an average of 3.91 (SD = 1.1) letters.

Participants who experienced verbal interference were more accurate (78.0%) in their interference memory than those who experienced spatial interference (66.1%), F(1,124) = 13.7, p = .0003. Interference memory accuracies for each interference type, training condition, and task condition are shown in Table 2.3. However, there were no other differences in interference accuracies based on congruency or training condition, or the interaction between these and interference type (all ps > .29).

**Table 2.2: Time Judgment Accuracies (%)**

<table>
<thead>
<tr>
<th>Learned earlier is up</th>
<th>No Interference</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>earlier key above</td>
<td>94.3</td>
<td>94.1</td>
<td>94.1</td>
</tr>
<tr>
<td>earlier key below</td>
<td>95.4</td>
<td>92.7</td>
<td>92.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learned earlier is down</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>earlier key above</td>
<td>94.6</td>
<td>93.2</td>
</tr>
<tr>
<td>earlier key below</td>
<td>94.6</td>
<td>93.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learned earlier is up</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>earlier key above</td>
<td>79.6</td>
<td>68.1</td>
</tr>
<tr>
<td>earlier key below</td>
<td>79.6</td>
<td>67.0</td>
</tr>
</tbody>
</table>

**Table 2.3: Interference Memory Accuracies (%)**

<table>
<thead>
<tr>
<th>Learned earlier is up</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>earlier key above</td>
<td>75.9</td>
<td>66.2</td>
</tr>
<tr>
<td>earlier key below</td>
<td>76.7</td>
<td>63.3</td>
</tr>
</tbody>
</table>
2.5 Experiment 1 Discussion

In this experiment, we asked whether learning a new way to talk about time can change the way people think about it. To do so, we taught participants new vertical spatio-temporal metaphors, and then probed their representations of time using a non-linguistic task that has demonstrated culturally consistent representations of time in prior work [43, 13, 44]. We asked further whether newly learned linguistic metaphors shape representations beyond simply teaching people new linguistic routines. To do so, we introduced verbal interference, using the same procedure that has eliminated people’s ability to engage in linguistic routines in prior work (e.g., [42, 127]).

We found that after learning the new ways to talk about time, participants showed metaphor-consistent representations in the non-linguistic task. That is, learning new metaphors in the lab induced the same kinds of differences between our two experimental groups as have been observed previously across cultures. Further, we found that the newly learned representations persevered even under conditions of verbal interference. These findings suggest that relational language, in this case new sets of metaphors, can indeed act as a formative force in creating new non-linguistic representations. It appears that participants who learned to say things like Tuesday is above Wednesday did more than simply learn a new way of talking. They activated representations of time on the vertical axis, representations that went beyond the specific linguistic forms and persisted even under conditions of verbal interference. When we encounter relational language, it appears we are invited to create new mental models. Language can invite us to imagine and represent the world in new ways.

Importantly, both types of interference (verbal and spatial) did, to some extent, interfere with performance. As would be expected, people were overall slower when performing the task under interference than with no interference. This helps assure
us that the interference paradigms were indeed engaging limited cognitive resources. What we do not find is a differential effect of either kind of interference across congruency conditions: people were not disproportionately slower under interference in the incongruent key-mapping condition as compared to the congruent condition. That is, while both interference paradigms made the task more difficult overall, neither kind of interference eliminated the specific mapping between space and time (e.g., up=earlier) being measured in the task.

If tying up linguistic or spatial working memory resources in general does not eliminate people’s specific space-time associations, what would? Our results suggest that what does affect the strength of specific space-time associations is exposure to new space-time associations. For example, throughout the course of a block (56 time judgments in these experiments), participants become increasingly acclimated to the response key assignments for that block. In effect, the current block’s key assignment acts as a new space-time association training, and the congruency effect gradually decreases.

This experiment shows that our representations of time are malleable and dynamic. For one, participants were able to learn a new system of metaphors for talking about time after just seeing five examples. They were able to apply the system to new sentences throughout the training portion of the experiment, but even more strikingly, the short 10-12 minute training session in the lab shaped performance on the subsequent non-linguistic time judgment task. People appear willing and able to learn new psychologically real spatializations for time through a variety of means (whether through verbal metaphor or through a pattern of motor movements to response keys).

Results of Experiment 1 show that we can use purely linguistic training to induce the same kinds of differences in thinking about time as have been found previously between cultures. These results are important for establishing that patterns in language, and specifically metaphor, can play a causal role in shaping people’s representations of
time. Experimentally changing what metaphors people use changes the way they think. These results suggest that learning new relational language can be a powerful tool in constructing new representations and expanding our cognitive repertoire. That is, we found both that language can *shape* thought, and that language can shape *thought*.

These results suggest that our representations of time are highly dynamic and can quickly change with new experience. Even patterns acquired over years of experience can be significantly pushed around by relevant new mappings. These findings also offer a context for thinking about the effects of linguistic training. Some theorists may find it surprising that a brief linguistic training in the lab could produce new patterns in non-linguistic behavior. These data suggest that instruction and experience of many types can shift our representations of time; even representations of a fundamental conceptual domain like time are more dynamic and less etched in stone than previously supposed.

2.5.1 **Open Questions**

The training paradigm used in Experiment 1 allows us the needed experimental control to make causal inferences. Of course, learning metaphors in the lab is very different from natural language experience. Do representations consistent with long-term natural language metaphors also persist despite verbal interference? I take up this question in Chapter 3 (Experiments 3 & 4).

A further question arises from the results of the spatial interference condition. In the current experiment, linguistically acquired space-time mappings also persisted under spatial interference from a block-grid interference task. Is this a special feature of space-time mappings that are acquired through language? Or would space-time mappings acquired through visuomotor experience likewise be unaffected by this kind of spatial interference (and verbal interference)? That is, are representations acquired through language necessarily different from those acquired through other means? I take up these
questions in Chapter 4 (Experiments 5-8).

2.6 Experiment 2 Introduction

The first experiment demonstrated that congruency effects between newly-learned metaphors and spatio-temporal biases were unaffected when people made time judgments under verbal (or spatial) interference. The congruency effects’ persistence under verbal interference suggests that the new representations were non-linguistic in nature. However, this conclusion rests on the assumption that the verbal interference used in Exp 1 would actually interfere with a representation that needs to be accessed through a linguistic routine.

Prior research gives weight to this assumption. For example, remembering a sequence of digits, similar to the task of remembering a sequence of consonants used here, eliminated language-consistency effects in a color discrimination task [127]. Without interference, Russian speakers were faster to differentiate between shades of blue with different lexical labels in Russian (siniy vs. goluboy) than shades with the same label. However, verbal interference eliminated that processing advantage for stimuli with different lexical labels. This finding suggests that the discrimination advantage Russian speakers originally demonstrated arose because they cued the linguistic labels in mind while making the distinctions. When they were unable to cue those labels because they were maintaining digits in mind, their language-consistent effect disappeared.

The exact verbal interference task used in this work has also interfered with another cognitive process that relies on a linguistic routine: counting dots [42]. Under verbal interference, participants were significantly less accurate to count dots that appeared on the screen than under spatial interference (also the same spatial interference paradigm used in this work). Again, these results suggest that because counting dots relies on an
internal linguistic routine (subvocally rehearsing digits: *one, two, three...*), maintaining consonants in working memory while counting hinders performance, while maintaining a sequence of boxes (spatial interference) does not hinder counting performance, as this second interference task does not tax the same verbal working memory resources as counting does. Thus, in prior work, the same verbal interference task that we used and a similar one have affected performance on linguistically-mediated tasks.

Despite the suggestions from prior work that our verbal interference paradigm is capable of interfering with linguistically-mediated representations, it is still possible that some feature of our population or experiment setup rendered the verbal interference task less effective at interfering with a linguistically-mediated representation than it did in prior work. To test this assumption, a separate group of UCSD participants completed the dot counting task used in prior work [42], while undergoing the same interference types used in that earlier work as well as in Exp. 1. If this verbal working memory task does indeed interfere with linguistically-mediated cognitive processes, participants should be less accurate when counting dots under verbal than spatial interference, as was found by Frank et al. [42].

### 2.7 Experiment 2 Methods

#### 2.7.1 Participants

Eighty undergraduates from UCSD took part in this experiment for course credit. They completed the experiment online.
2.7.2 Materials & Procedure

The experiment was programmed in jsPsych [33] so it could be deployed online and executed in a browser. Participants were randomly assigned to either the verbal or spatial interference condition. Interference was calibrated as in Exp. 1, using a 2-up-1-down staircase method, in which participants remembered the verbal or spatial distractors while completing a visual search task, and then responded with the distractors they had kept in mind.

After interference calibration, each participant’s threshold number of distractor items was used for the main counting task. For this task, participants first encoded the verbal or spatial interference (letters or boxes appeared on the screen one at a time), and then they counted dots that flashed on the screen one at a time. Next they reported the number of dots they saw, by typing in a digit in a textbox that appeared, and finally repeated the interference distractors they had seen. As in the precedent work [42], dots were shown for 500 ms each, with a 250 ms interval between presentations. There were 6 trials showing each of the quantities between 1-10, and there was one trial showing each of the quantities from 11–15, to reduce the possibility that participants would notice a ceiling on the quantity distribution.

2.7.3 Participants & Trial Inclusion Criteria

All participants were included. Only trials for which the interference accuracy was correct (65.8% of all trials) were analyzed. If participants failed to even encode the interference distractors, then they would have essentially been completing the dot counting task without interference, and therefore would not contribute to an understanding of how well people count under these two types of interference. Although the percentage of included trials may seem low, we actually had more dot counting trials remaining after
eliminating incorrect interference trials than remained in the work by Frank et al. ([42]; in which interference memory was correct only for 52.5% of trials).

2.7.4 Analyses

We conducted the same analyses as in Frank et al. [42] to ensure that the current results replicated those earlier ones. Thus, we fit a generalized linear mixed model to correct trials on cardinalities 1-10 including effects of interference condition and target quantity, plus their interaction, and a random effect of participants.

2.8 Experiment 2 Results

The number of distractors that participants saw in the spatial interference condition while counting dots (M = 4.7, SE = 0.22) did not differ from the number of distractors participants saw in the verbal interference condition (M = 4.4, SE = 0.15; t(69) = 0.28, p = .78).

The generalized linear mixed model did not show an effect of interference condition (b = -0.03, SE = 0.34, p = .93) or of target quantity (b = -0.03, SE = 0.03, p = .38), but there was an interaction between interference type and quantity (b = -0.13, SE = 0.04, p = .0008). These results nearly replicate those found by Frank et al. [42]. The only difference is that the earlier work did find an effect of target quantity (people were more accurate for smaller quantities), but the inferentially crucial effect in both the prior work and this replication is the interaction between interference type and quantity. This interaction demonstrates that the relationship between target quantity and accuracy was different by condition. Specifically, in the spatial interference condition, participants were able to count fairly accurately, even for the larger target quantities. In the verbal interference condition, though, participants had to estimate more, which especially impacted their
accuracy as the target quantity increased. Figure 2.4 shows these results.

![Figure 2.4](image)

**Figure 2.4**: Probability distribution over response quantities for each quantity tested. (darker = higher probability).

Reaction time results were consistent with the accuracy results: people were numerically slower to respond to the dot counting task under verbal interference (M = 3685ms, SE = 164) compared to spatial (M = 3183ms, SE = 123). Although this difference was not significant (p = .88), there was a marginal interaction between interference condition and target quantity that mirrored the counting accuracy results: as the target quantity increased, reaction times under verbal interference disproportionately suffered (b =87.4, SE = 50.3, p = .08). These results suggest that there was not a speed-accuracy tradeoff: participants were not more accurate under increasing target quantities in the spatial interference condition because they took more time to respond.

### 2.8.1 Comparison to Experiment 1

Since the purpose of this experiment was to demonstrate that the verbal dual task used in Exp 1 does interfere with a linguistic routine, we compared the number
of distractors participants in the verbal interference experienced in this experiment to those experienced in Exp 1. On average, participants saw more verbal distractors in the current experiment (M = 4.4, SE = 0.10) than in Exp 1 (M = 4.0, SE = 0.10; t(78) = 2.09, p = .04), so we eliminated participants from the current experiment to arrive at an overall mean of 4 verbal distractors (and 4 spatial distractors) and repeated the crucial analysis, to ensure that the interaction between interference condition and target quantity on counting accuracy remained once the number of distractors was equated to those in Exp 1 (b =-0.13, SE=0.05, p = .003).

2.9 Experiment 2 Discussion

This replication reaffirms that the verbal interference used in Experiment 1 (and subsequent experiments) has a different effect on quantity encoding than the matched spatial interference. If a cognitive process (such as counting) relies on a linguistic routine, this verbal interference paradigm should affect performance on that task more than the spatial interference paradigm does. This finding strengthens the conclusions from Experiment 1, that newly-learned spatio-temporal metaphors create mental representations that are non-linguistic.

Chapter 2, in part, is a reprint of the material as it appears in Hendricks, R. K. & Boroditsky, L. (2017). New Space-Time Metaphors Foster New Nonlinguistic Representations, *Topics in Cognitive Science*. The dissertation author was the primary investigator and author of this paper.
Chapter 3

Chinese speakers’ space-time associations

3.1 Abstract

What is the nature of the language-consistent mental representations observed in English-speaking participants who learned a new system of metaphors in the lab? The results of Experiment 1 suggested that they do not necessarily remain linguistic in nature, since they were unaffected by verbal interference. Experiment 3 seeks to substantiate these findings with a more ecologically valid case. As such, we examine the nature of native Chinese speakers’ mental space-time associations by subjecting them to the same verbal and spatial dual task used in Experiment 1. We find that natural language congruency effects (placing earlier events above later, as in Chinese linguistic metaphors) persist under interference, suggesting that they do not require verbal working memory in the moment to manifest. Following up on those findings, in Experiment 4 we ask: Would the same natural language-consistent results hold if we could rule out all confounds, ensuring that these representations arose specifically from experience in language and
not some other influence? To test that, we train Chinese-English bilinguals to use new metaphors in the lab, as in Experiment 1, thus combining crucial features of Experiment 1 (training participants to use new systems of metaphor in the lab) and Experiment 3 (examining Chinese-English bilinguals’ mental representations). We find that mental representations consistent with long-term and newly-learned metaphors are comparable in size and nature.

3.2 Introduction

The world’s languages differ from one another in a myriad of ways, including lexicon, grammar, and systems of metaphors. Prior work has shown that speakers of different languages also often think differently in ways that are consistent with patterns in their languages (for review, see [133]). In other words, people tend to hold language-consistent mental representations; that is, the way they think is consistent with the way they typically talk. What is the nature of these language-consistent representations? One possibility is that when we learn something through language, we create an online linguistic routine. A linguistic representation may influence behavior by participating alongside other representations. Some researchers have suggested that this is the only way that language influences cognition; they suggest that language cannot shape nonlinguistic representations (e.g., [90]). However, it is also possible that through experience with language one can acquire or build new non-linguistic representations (e.g., spatial, visual, olfactory, symbolic, etc).

One method that prior research has used to distinguish linguistic and non-linguistic representations is to have people engage in a verbal task while doing another task of interest. The logic of this paradigm is that cognitive interference occurs if the main task requires the verbal working memory resources that are also engaged in the verbal
dual task. In some cases, verbal interference eliminates congruency effects between language and thought. For example, Russian makes an obligatory distinction between light blues (goluboy) and dark blues (siniy), that in English can be referred to with the single term blue [127]. Accordingly, in one experiment Russian speakers were faster to distinguish shades of blue with different linguistic labels than the same label, while English speakers showed no such advantage. Under verbal interference, however, Russian speakers no longer showed this language-consistent pattern in their color judgments.

Here we investigate the nature of language-consistent mental representations in the example domain of time; specifically, people’s representations of temporal relations. To talk about temporal relations people often use spatial language [28, 74]. For example, English metaphors commonly refer to events in the future as ahead or in front and the past as behind.

People don't just talk about time using space; they also think about time using space. For example, prior work has shown that English speakers show space-time congruency effects, with faster responses for future events in front and past events behind [117, 104, 121, 102]. They also lean forward when talking about the future and lean back when talking about the past [86] and gesture forward when talking about the future and backward when talking about the past [20]. Across these studies, language-consistent spatial representations manifest in the absence of spatial language during the task.

The specific ways in which time is mapped to space differ considerably across languages. For example, while English relies primarily on horizontal (front/back) spatial terms for talking about time, Mandarin commonly relies both on horizontal and on vertical (up/down) terms. In Mandarin, earlier events are said to be up and later events are down. Although English does occasionally use vertical metaphors (like passing knowledge down), these metaphors are not as systematic, productive, or frequent as they are in Mandarin [103]. This difference between Mandarin and English has provided
opportunities to investigate corresponding differences in thought.

Consistent with linguistic tendencies, Mandarin speakers also show vertical representations of time that place earlier events above later ones. Research has revealed these representations when people arrange image sequences [87, 131], gesture while talking about temporal relations [58], and make time judgments in non-linguistic tasks [13, 10, 44].

For example, in tasks that use pictures as stimuli and spatial keys as responses Mandarin speakers are more likely to show an association between earlier and above than English speakers [87, 13, 44]. In fact, while Mandarin speakers consistently represent earlier events as above later ones, English speakers do not. In the task used by Boroditsky et al. [13] and Fuhrman et al. [44], which was adapted for this current work, participants see an initial image (e.g., Julia Roberts in her 20s) followed by a second image (e.g., either a younger or older Julia). Participants determine whether the second image shows a conceptually earlier or later time than the first, and they press a key to respond. People are faster to respond when doing so is congruent with their mental representation than incongruent.

Thus, prior work using this task has revealed language-consistent differences in thought. What is the nature of the representations that give rise to these differences in behavior? Have English and Mandarin speakers acquired different linguistic routines for the purpose of ordering events, for example reciting earlier is up as they do the task? Or have they acquired different non-linguistic ways of representing time? While the task itself does not require the use of language, it is of course possible that people might privately call on a verbal routine while doing the task. If it was necessary for people to engage internal linguistic routines to show language-consistent behavior, it should be possible to disrupt language-consistent behavior by limiting peoples ability to engage in linguistic routines.
To test this, Hendricks and Boroditsky (Ch. 1, this dissertation; [65]) used verbal interference while people made the same time judgments as in prior work [13, 44]. Before the task, English-speaking participants learned a system of metaphors that placed earlier events either above or below later ones (i.e., *breakfast is above dinner* or *breakfast is below dinner*). Participants responded more quickly when the response arrangement was congruent with their newly learned metaphors than incongruent. In other words, learning new spatial metaphors to talk about time created new ways of thinking about time. Importantly, the congruency effects persisted under verbal interference, suggesting that these participants’ new representations of time were non-linguistic, persisting even when verbal working memory resources were occupied.

In this chapter we ask whether representations consistent with natural language metaphors also persist despite verbal interference. Further, we directly compare results arising from newly learned and natural language metaphors in the same group of participants.

In Experiment 3, Mandarin-English bilinguals completed the same time judgments with the same interference conditions as in earlier work (Ch. 1, this dissertation; [65]). One group of participants consisted only of Chinese-English bilinguals, and therefore had long-term experience with metaphors that place earlier events above later ones. This group was randomly assigned to participate either in Mandarin (referred to throughout as the Chinese-in-Mandarin group) or in English (referred to as the Chinese-in-English group) to further probe for the strength and nature of language-consistent representations when the language of interest is or is not being used in the current context. As a control, a final group of participants had no Chinese background (referred to as the English group).

In Experiment 3, we found natural language-consistent congruency effects among Chinese-English bilinguals: these participants were faster to respond when the earlier
response was above the later. However, this finding cannot rule out the possibility that some extralinguistic factor (e.g., writing direction), rather than linguistic metaphors, drove the congruency effect [120, 8].

To investigate whether congruency effects like the ones observed in Experiment 3 can truly come from experience with language, participants must be randomly assigned to a language condition, which is not possible for natural languages. In Experiment 4, we trained Chinese-English bilinguals to use vertical metaphors to talk about time in English. They learned that earlier events were either above later ones (congruent with their natural language metaphors) or below later ones (incongruent) and completed the same time judgments under the same conditions of interference as in Experiments 1 and 3.

3.3 Experiment 3 Methods

3.3.1 Participants & Trial Inclusion

Overall, 312 undergraduates at UC San Diego participated for course credit. As in Experiment 1, participants were excluded if their time judgment or interference accuracy for at least one block was below 3 standard deviations from the group mean judgment or interference accuracy. Exclusions for each of the groups (Chinese-in-Mandarin, Chinese-in-English, English) were made separately. Five participants were excluded from the Chinese-in-Mandarin group (4 for poor judgment accuracy and one for poor interference accuracy). Six participants were excluded from the Chinese-in-English condition, all for poor task performance, and 8 were excluded from the English condition, also for poor task performance. We recruited replacements for all excluded participants to arrive at a fully-counterbalanced sample within each condition: for the Chinese-in-Mandarin group, this left 17 participants per cell (interference condition and block order), and for the other two groups, there were 16 participants per cell.
On a 1-5 scale, participants in the Chinese-in-Mandarin group reported an average Chinese proficiency of 4.9/5 (SD = 0.2) and an average English proficiency of 4.1/5 (SD = 0.7). Those in the Chinese-in-English group reported an average Chinese proficiency of 4.6/5 (SD = 0.9) and an average English proficiency of 4.1/5 (SD = 0.8). The Chinese-in-Mandarin group reported greater Chinese proficiency than the Chinese-in-English group (t(201) = 3.35, p = .001), although there was no difference in how these participants were recruited. For this reason, we took Chinese proficiency into account when examining the effect of natural language experience on vertical space-time associations by including self-rated proficiency ratings as predictors in the regression models.

As in Experiment 1, only trials for which the time judgment was correct (92.6%) as well as interference memory (80.5% of all correct judgments) were included. Further, we analyzed only judgments for which the reaction time was within 3 standard deviations of the participant’s cell mean and greater than 250ms (97.8% of all correct trials).

### 3.3.2 Materials & Procedure

Participants in the Chinese-in-Mandarin group were greeted and instructed in Mandarin by a native Mandarin speaker. All others were greeted and instructed in English (by the same research assistants who ran the experiment for the Chinese-in-Mandarin group). Regardless of the language being spoken in the experiment, response keys were never referred to with spatial labels; instead, they were always referred to by their color (one was red and the other was blue).

**Time Judgments**

We used the same non-linguistic space-time implicit association task described earlier that has revealed cross-linguistic differences in temporal representations for English, Mandarin, and Hebrew speakers [13, 44, 43] and for English speakers taught to
use new temporal metaphors [65].

**Interference**

Participants completed the time judgments under verbal, spatial, or no interference, following the same procedures as in Hendricks and Boroditsky [65]. The only adaptation we made to these methods was using digits in the verbal interference condition (instead of consonants) so that the distractors would not be inherently English focused, but could be translated to Chinese for the Chinese-in-Mandarin group. For the interference groups, participants were asked to remember either a string of digits (verbal) or a pattern of boxes (spatial) while performing the time judgments. As in prior work, these participants completed a calibration block at the beginning of the experiment to ensure that the interference difficulty was appropriately scaled across individuals (for details, see Experiment 1 [Chapter 2]).

After interference calibration, each participant completed two blocks of time judgments. The two time judgment blocks had opposite key response assignments: in one block, the *earlier* key was above *later* and in the other the *earlier* key was below *later*. Block order was counterbalanced across participants. Keys were marked with colors and were not labeled with words.

### 3.3.3 Analyses

We fit the data with mixed effects models using Laplace Approximation using the *lmer()* function within the *lme4* analysis package in R [7]. The models included time judgment condition (earlier key above or below later), interference condition, and participant group (forward-difference coded so that Chinese-in-Mandarin > Chinese-in-English > English) as fully-crossed fixed effects, and subjects and items as random effects. The model had random intercepts and slopes for the within-subjects variable of
time judgment condition.

3.4 Experiment 3 Results

3.4.1 Time Judgment Reaction Times

Did participants show earlier-is-up biases?

If participants think of earlier events as above later ones, consistent with Chinese natural language metaphors, they should be faster to make judgments when the *earlier* key is above the *later* than the reverse. Collapsing across all participants, reaction times were faster when the *earlier* key was above the *later* key (M = 1808 ms) than when the *later* key was above (M = 1865 ms), as indicated by a main effect of time judgment condition (b = -58.7 ± 28.2 SEM, t = -2.08, p = .04).

Further, there was an interaction between time judgment condition and group (English speakers vs. Chinese-in-English), demonstrating that the early-up bias was stronger for Chinese speakers (participating in English) than English speakers (b = -101.1 ± 29.7 SEM, t = -3.4, p = .0007). Early-up biases did not differ between the two groups of Chinese speakers (b = 9.8 ± 56.9 SEM, t = 0.2, p = .86). These results suggest that experience with Chinese natural language metaphors is the crucial factor for developing a metaphor-consistent early-up bias; it is not necessary for Chinese to be spoken in the moment for the bias to manifest in behavior. Reaction times by participant group and interference condition are shown in Table 3.1, and early-up biases by participant group and interference condition are shown in Figure 3.1.

We further explored the relationship between features of a participant’s language background (their self-reported Chinese language proficiency) and their early-up bias. We did not find that greater proficiency in Chinese was correlated with a greater early-up
bias ($r = -0.02, n = 208, p = .76$), likely because participants overall reported being very proficient in Chinese ($M = 4.88/5$), with not much variation in responses ($SD = 0.48$).

**Table 3.1:** Reaction Times (SD) for Chinese-English bilinguals and English speakers who made temporal judgments (without metaphor training) by interference type. All reaction times are in ms.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chinese-in-Mandarin</strong></td>
<td>earlier key above 1955 (448)</td>
<td>1832 (342)</td>
<td>1823 (535)</td>
</tr>
<tr>
<td></td>
<td>earlier key below 1920 (514)</td>
<td>1955 (457)</td>
<td>1991 (523)</td>
</tr>
<tr>
<td><strong>Chinese-in-English</strong></td>
<td>earlier key above 1546 (408)</td>
<td>1872 (463)</td>
<td>1648 (291)</td>
</tr>
<tr>
<td></td>
<td>earlier key below 16343 (494)</td>
<td>1944 (624)</td>
<td>1799 (435)</td>
</tr>
<tr>
<td><strong>English</strong></td>
<td>earlier key above 1777 (525)</td>
<td>1898 (468)</td>
<td>1929 (944)</td>
</tr>
<tr>
<td></td>
<td>earlier key below 1902 (509)</td>
<td>1853 (349)</td>
<td>1770 (551)</td>
</tr>
</tbody>
</table>

**Figure 3.1:** Results of Experiment 3: Early-up biases were calculated by subtracting each participant’s mean reaction time when the earlier key was above the later (Mandarin metaphor-congruent) from the participants mean reaction time for the opposite, Mandarin metaphor-incongruent condition.

**Did verbal interference affect the earlier-is-up biases?**

Overall, people were not faster without interference compared to under spatial ($b = -97.1 \pm 58.9$ SEM, $t = -1.6, p = .10$) or verbal ($b = 42.3 \pm 52.8$, $t = 0.8, p = .42$)
However, our primary goal was to investigate the size of congruency effects under verbal interference, rather than overall reaction times to investigate whether natural language-consistent mental representations of time are similar in nature to those consistent with nearly-learned metaphors. Was the magnitude of the early-up bias different under verbal interference than under spatial? This analysis included only the two groups with Chinese language experience, since it is unclear how we would interpret an effect of interference on English speakers early-up biases, as there was no such systematic bias.

Among Chinese speakers (Chinese-in-Mandarin and Chinese-in-English), there was not an interaction between interference type (verbal vs. spatial) and task condition (b = -61.1 70.4 SEM, t = -0.9, p = .39). There was also no interaction between interference type (verbal vs. none) and task condition (b = 65.5 68.6 SEM, t = 1.0, p = .34), though the early-up bias was marginally greater under spatial interference than under no interference (b = 126.6 69.0 SEM, t = 1.83, p = .07). It is difficult to interpret this interaction between one group with a dual task and one group without because of the fundamentally differing natures of the tasks, but this marginal effect may be a subtle suggestion about the involvement of spatial working memory resources in the moment when a language-consistent temporal representation is evident in behavior. However, the focus of this work is on whether such representations must remain linguistic in nature. The lack of difference in participants’ early-up biases under verbal as compared to spatial interference (plus verbal compared to none) indicates that as in Experiment 1, language-consistent space-time associations do not seem to require linguistic working memory resources in the moment when they manifest in behavior.

Additionally, we verified that the interference types did not affect the early-up bias differently for the two groups (Chinese-in-Mandarin vs. Chinese-in-English; verbal
vs. spatial interference: $b = 40.4 \pm 140.8$ SEM, $t = 0.3$, $p = .77$; verbal vs. no interference: $b = 179.0 \pm 137.3$ SEM, $t = 1.3$, $p = .19$; spatial vs. no interference: $b = 138.6 \pm 137.9$ SEM, $t = 1.0$, $p = .32$). In sum, the metaphor-consistent early-up bias was not different under the two interference types, suggesting that it does not require either verbal or spatial working memory resources (as taxed by these traditional working memory tasks) to manifest.

One puzzle in the data is that post-hoc tests revealed that the group of Chinese-English bilinguals who completed the experiment in Chinese and without interference did not show the native language-consistent earlier-is-up bias ($b = 32.7 \pm 77.1$ SEM, $t = 0.4$, $p = .67$) that has been documented in other work [13, 44]. A potential explanation for this lack of effect may be that by chance those in the Chinese-in-Mandarin group who were assigned to the no interference group did not have as strong of Chinese backgrounds as those assigned to the other conditions. However, this was not the case, since among Chinese-in-Mandarin participants, those in the no interference group had an average proficiency of 4.93/5 (SD = 0.25), which was not different from the mean Chinese proficiency in the verbal interference group ($M = 4.96$, SD = 0.21; $t(50.7) = 0.4$, $p = .72$) nor the mean Chinese proficiency in the spatial interference group ($M = 4.97$, SD = 0.18; $t(51.4) = 0.63$, $p = .53$).

Another potential explanation to the lack of natural language congruency effect in the Chinese-in-Mandarin participants without interference might lie in the speed with which they completed the time judgments. Although they were not significantly slower than Chinese-in-Mandarin participants who did the task under spatial interference ($M = 1938$ (SD = 556) ms vs. $M = 1917$ (SD = 509) ms; $t(47.2) = 0.1$, $p = .92$) or those who did the task under verbal interference ($M = 1950$ (SD = 311) ms; $t(59) = 0.1$, $p = .88$), they were significantly slower than Chinese-in-English participants without interference ($M = 1595$ (SD = 428) ms $t(54.4) = 2.7$, $p = .01$). Although this difference does not clearly explain why we did not observe a congruency effect for the Chinese-
in-Mandarin group without interference, it does point to one way in which this group was not necessarily representative of the larger population of Chinese-English bilinguals. Perhaps, for example, they were thrown off by completing the experiment in Mandarin, and thus were slower for the task than they would have been otherwise. Participants in the Chinese-in-Mandarin group who were in the verbal and spatial interference conditions might have also been thrown off by the Mandarin experiment, but they had a block of interference calibration first in which to become acquainted with using Mandarin in the experiment setting, thus leaving their time judgment reaction tasks unaffected.

3.4.2 Time Judgment Accuracies

Overall, participants were more accurate when the earlier response key was above the later than the reverse ($b = -0.26 \pm 0.08$ SEM, $z = -3.3$, $p = .001$). Detailed judgment accuracies are shown in Table 3.2.

As might be expected, participants made more accurate judgments when they weren’t under interference (93.4%) than when they were under either verbal (92.6%; $b = -0.28 \pm 0.12$ SEM, $z = -2.3$, $p = .02$) or spatial (92.0%; $b = -0.38 \pm 0.12$ SEM, $z = -3.1$, $p = .002$) interference. However, there was no difference in judgment accuracies for participants under spatial vs. verbal interference ($b = -0.01 \pm 0.10$ SEM, $z = -0.1$, $p = .95$).

The early-up bias evident in time judgment accuracies did not differ by participant group (Chinese-in-Mandarin vs. Chinese-in-English: $b = 0.05 \pm 0.19$ SEM, $z = 0.3$, $p = .80$; Chinese-in-English vs. English: $b = 0.08 \pm 0.19$ SEM, $z = 0.4$, $p = .70$).

3.4.3 Interference Accuracies

Participants under verbal interference were more accurate (72.4%) than those under spatial (68.1%; $b = 1.22 \pm 0.14$ SEM, $z = 8.9$, $p < .0001$), consistent with prior
Table 3.2: Time judgment accuracies (%) by interference type

<table>
<thead>
<tr>
<th></th>
<th>earlier key above</th>
<th>earlier key below</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese-in-Mandarin</td>
<td>93.1, 94.2</td>
<td>92.5, 91.5</td>
</tr>
<tr>
<td></td>
<td>92.6</td>
<td>90.0</td>
</tr>
<tr>
<td>Chinese-in-English</td>
<td>92.5, 94.0</td>
<td>93.1, 92.6</td>
</tr>
<tr>
<td></td>
<td>91.7</td>
<td>92.8</td>
</tr>
<tr>
<td>English</td>
<td>92.3, 94.2</td>
<td>92.7, 92.8</td>
</tr>
<tr>
<td></td>
<td>92.8</td>
<td>92.3</td>
</tr>
</tbody>
</table>

work using these same interference paradigms [42]. At the same time, participants in the verbal interference condition saw more distractors on average (M = 5.7) than those in the spatial interference condition (M = 4.5). These findings demonstrate the importance of calibrating the interference task for each individual as we did here, since it may not be possible for the spatial and verbal tasks to be equated for difficulty. Table 3.3 shows interference accuracies by participant group, interference type, and task condition.

We next examined whether interference accuracy differed based on the orientation of the response keys. Participants were more accurate at recalling the interference when the accompanying trial was congruent with Chinese natural language metaphors – when the earlier key was above the later – than the reverse (b = 0.12 0.05 SEM, z = 2.6, p = .01). This early-up bias evident in interference accuracies did not differ across participant groups (Chinese-in-Mandarin vs. Chinese-in-English: b = -0.01 0.16 SEM, z = -0.1, p = .94; Chinese-in-English vs. English: b = 0.27 16 SEM, z = 1.6, p = .10). However, this pattern in accuracies was only evident for participants under spatial interference, as demonstrated by an interaction between task condition and interference type (b = -0.24 .07 SEM, z = -3.6, p = .0003).
Table 3.3: Interference Accuracies (%)

<table>
<thead>
<tr>
<th>Language</th>
<th>Interference Type</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese-in-Mandarin</td>
<td>earlier key above</td>
<td>73.3</td>
<td>64.5</td>
</tr>
<tr>
<td></td>
<td>earlier key below</td>
<td>73.9</td>
<td>61.6</td>
</tr>
<tr>
<td>Chinese-in-English</td>
<td>earlier key above</td>
<td>68.2</td>
<td>73.7</td>
</tr>
<tr>
<td></td>
<td>earlier key below</td>
<td>69.5</td>
<td>73.1</td>
</tr>
<tr>
<td>English</td>
<td>earlier key above</td>
<td>72.9</td>
<td>69.3</td>
</tr>
<tr>
<td></td>
<td>earlier key below</td>
<td>77.6</td>
<td>66.9</td>
</tr>
</tbody>
</table>

3.4.4 Comparison to Experiment 1, Metaphor Training

Does the magnitude of participants’ language-consistent space-time associations differ depending on whether these associations are consistent with newly-learned metaphors in the lab (Experiment 1) or long-term metaphor exposure through natural language (this experiment)? To test this, we used a mixed effects model that included experiment (1 or 3), congruency (congruent or incongruent, with respect to newly-learned metaphors for Exp 1 participants and with respect to long-term Chinese metaphors for Exp 3 participants), and interference type as fully-crossed fixed effects, and participants and items as random effects. This analysis revealed that the magnitude of the language-consistent congruency effect did not differ across experiment, since there was not an interaction between experiment and congruency (b = 30.4 ± 24.7 SEM, t = 1.2, p = .22).

Further, the effect of interference on congruency effects did not differ across experiment, as there were no 3-way interactions among experiment, congruency, and interference type (verbal vs. spatial interference: b = 57.6 ± 67.2, t = 0.9, p = .39; verbal vs. no interference: b = 23.2 ± 42.3 SEM, t = 0.5, p = .58; spatial vs. no interference: b = 11.2 ± 42.0 SEM, t = 0.3, p = .79). This result suggests that at least in terms of persistence under interference, Chinese speakers’ natural language metaphor-consistent mental representations are similar in nature to the representations created by teaching participants new spatio-temporal metaphor systems in the lab.
### 3.5 Experiment 3 Discussion

What is the nature of language-consistent mental representations? This experiment examined the nature of Chinese-English bilinguals’ vertical representations of time. We built on prior work showing that Mandarin speakers have vertical mental timelines with earlier events above later ones [13, 44] and work showing English speakers who learn new temporal metaphors show metaphor-consistent mental representations of time even under verbal interference (Experiment 1, this dissertation [65]).

In the current experiment, participants made earlier/later judgments about sequences of events with vertically arranged response keys. We first replicated prior work showing that Chinese speakers showed representations consistent with their natural language metaphors (earlier is above later [13, 44]). Those representations persisted under verbal interference, suggesting that they do not rely on verbal routines.

Of course, we cannot be certain that the language-consistent patterns we observed for Mandarin speakers in Experiment 3 actually arose from experience with language, as any number of other factors might have given rise to the same patterns. Would the same results hold if we could rule out potential confounds and randomly assign participants to groups? In Experiment 4, we trained Chinese-English bilinguals to use vertical metaphors to talk about time. Can we simultaneously observe evidence of mental representations consistent with long-term natural language metaphors and newly-learned linguistic metaphors?

**Experiment 4: Chinese-English Bilinguals Learn New Linguistic Metaphors**
3.6 Experiment 4 Methods

3.6.1 Participants & Trial Inclusion

Fifty-eight undergraduates at UCSD participated for course credit. They were recruited as part of prior work investigating the effects of learning a new metaphor (Experiment 1, this dissertation [65]. Our advertisement only specified that participants be native English speakers, but many of the participants who identified themselves as native English speakers turned out to also have Chinese language experience (as detected in a post-experimental language experience questionnaire). This allowed us a serendipitous sample of Chinese-English bilinguals. These participants were not recruited for their Chinese language experience and had no idea that this aspect of their experience might be of interest to us. The experimenters likewise did not know the participants had Chinese language experience until after the task was completed. These participants reported an average English proficiency of 4.72 (SD=.26) and a Chinese proficiency of 3.16 (SD=1.18).

Participant and trial exclusion criteria were the same as in the prior experiments. We eliminated 2 participants for low accuracy learning the new metaphors, 5 for low time judgment accuracies, and 3 for low interference accuracies. After exclusions, we analyzed data from a fully counterbalanced sample of 48 participants. We included only correct time judgments (94%) with correct interference (72%) and reaction times within 3 SDs of each participant’s cell mean (98%).

3.6.2 Materials & Procedure

After interference calibration (for those in either the spatial or verbal interference conditions), participants learned a new way to talk about time. This procedure was the same as in Experiment 1 (Ch 2, this dissertation). Participants read five example
sentences showcasing this new system. For half of the participants, earlier events were said to be above or higher and later events were said to be below or lower. This system was reversed for the other half of participants so that earlier events were below/lower and later events were above/higher. Participants completed 90 test trials to practice this new system.

After training, participants began the first time judgment block accompanied by interference when applicable. From here, the rest of the experiment was identical to the procedure in the prior experiments (Exp 1, Ch 2 & Exp 3, this chapter). Which metaphor system people were trained on and whether the first time judgment block was congruent or incongruent with the training were counterbalanced across participants. The entire experiment was conducted in English.

The analyses were identical to those used in Experiment 3 with the addition of training condition as a fully-crossed fixed effect and without the effect of participant group.

## 3.7 Experiment 4 Results

### 3.7.1 Time Judgment Reaction Times

Participants were faster when the time judgment responses were congruent with their newly learned metaphors (1563 ms) than when the responses were incongruent (1660 ms; $b = 99.62$, $SE = 41.88$, $p = .02$). Table 3.4 shows mean reaction times. Participants were also faster when the earlier response was above the later response (1551 ms) than for the reverse (1672 ms; $b = 124.79$, $SE = 42.47$, $p = .005$, consistent with Mandarin metaphors, the prior experiment, and with prior work [44]. That is, participants responded consistently with both their native language and newly-learned metaphors.
Table 3.4: Time judgment reaction times in ms (SD) by Interference Type.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learned earlier is up</td>
<td>1393 (260)</td>
<td>1596 (273)</td>
<td>1494 (495)</td>
</tr>
<tr>
<td></td>
<td>1533 (410)</td>
<td>1953 (367)</td>
<td>1650 (442)</td>
</tr>
<tr>
<td>Learned earlier is down</td>
<td>1376 (312)</td>
<td>1722 (312)</td>
<td>1724 (321)</td>
</tr>
<tr>
<td></td>
<td>1389 (231)</td>
<td>1740 (215)</td>
<td>1767 (321)</td>
</tr>
</tbody>
</table>

Reaction times did not differ for participants who underwent verbal (1670 ms) as opposed to spatial interference (1677 ms; b = 6.30, SE=40.58, p = .88) or between those who completed the task under verbal and no interference (1387 ms; b = 83.0, SE=48.8, p = .10). However, the spatial interference group was slower than the no interference group (b = 105.0, SE=49.7, p = .04). The size of the congruency effect with respect to the newly-learned metaphors did not differ across the interference groups (ps > .23). This finding mirrors prior work showing that English speakers tested on this same task also show representations consistent with newly learned metaphors that persist under verbal interference (Experiment 1, this dissertation[65]). The Chinese language-consistent congruency effect also did not differ across the interference groups (all ps > .18).

We also compared the size of the congruency effects across all participants with respect to their newly learned metaphor and with respect to Chinese natural language metaphors. A 2 (congruency: with respect to metaphor training or natural language) x 3 (interference type) ANOVA on each participant’s congruency effects revealed no difference in the size of congruency effects consistent with new metaphors and natural language, F(1,45) = .14, p = .72. These results are shown in Figure 3.2.

3.7.2 Proficiency & Congruency Effects

As in Experiment 3, we calculated an overall vertical bias for representing earlier events as above later ones for each participant. More proficient Chinese speakers showed a stronger vertical bias consistent with Chinese metaphors, reflected by a positive correa-
Figure 3.2: Results of Experiment 4: Congruency effects for Chinese-English bilinguals with respect to either newly-learned metaphors (black) or natural language metaphors (light gray).

Correlation between proficiency and biases placing earlier events above later ones \( r = .30, n = 44, p = .05 \). It is likely that we found a correlation between proficiency and natural language congruency effects here (and not in Exp. 3) because there was greater variation in participants’ Chinese proficiencies in this experiment (\( M = 3.16; SD = 1.18 \)) than in the prior (\( M = 4.88; SD = 0.48 \)). This difference likely arose because in Exp. 3, our ad for participants specified that they must be able to speak Chinese, leading to overall more proficient Chinese speakers who participated in Exp 3 as compared to Exp. 4, in which our ad specified that participants should self-identify as a native English speaker.

### 3.7.3 Accuracies

The group who learned that earlier events are above later ones (93.6%) and those who learned that earlier events are below (94.0%) did not differ in their accuracy for learning the new metaphors \( b = .05, SE = .24, p = .84 \).

There was also no difference in accuracy for time judgments congruent (95%)
or incongruent (94%) with newly learned metaphors (b = .33, SE=.38, p = .38). There was similarly no difference in accuracy for time judgments that were congruent (95%) or incongruent (94%) with Chinese natural language metaphors (b = .48, SE=.31, p = .12). Time judgment accuracies are shown in Table 3.5.

<table>
<thead>
<tr>
<th>Table 3.5: Time judgment accuracies (%) by interference type, training, and task conditions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
</tr>
<tr>
<td>learned earlier is up</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>learned earlier is down</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

There were no differences in interference accuracy between those who experienced spatial (66%) and verbal interference (74%; b = 1.04, SE=.68, p = .13). There was also no difference in interference accuracy for judgments congruent (70%) or incongruent (70%) with the newly-learned metaphors (b = .07, SE=.35, p = .84) or those congruent (72%) or incongruent (68%) with natural language metaphors (b=.19, SE=.25, p = .45). Interference accuracies are shown in Table 3.6.

<table>
<thead>
<tr>
<th>Table 3.6: Interference accuracies (%).</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbal</td>
</tr>
<tr>
<td>learned earlier is up</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>learned earlier is down</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Participants in the spatial interference group completed the experiment with an average of 4.0 (SD = 1.3) boxes, and those in the verbal interference group had an average of 4.25 (SD = 1.2) letters, which did not differ for the two groups (p = .58).
3.7.4 Comparison to Experiment 3, Natural Language Congruency Effects

Do Chinese-English bilinguals trained on a new metaphor system that is consistent with the one in their natural language show stronger early-up biases than those who are not trained on any new system of metaphors in the lab? If the bilinguals are trained on a new natural language-inconsistent system, do they show smaller early-up biases than they would otherwise?

To address these questions, we constructed a fully-crossed mixed effects model that included time judgment condition (earlier key is above or below later), interference condition, and participant group (whether trained on natural language-consistent, not trained in the lab, or trained on natural language-inconsistent metaphors) as a forward-difference coded factor.

Overall, there was a marginal main effect of task condition: participants were faster when the earlier response was above the later response (M = 1758 ms; b = -67.9 35.9 SEM, t = -1.9, p = .06) than the reverse (M = 1853 ms). The magnitude of this bias differed by participant group (for those trained on consistent metaphors vs. untrained, b = -164.8 53.8 SEM, t = -3.1, p = .002; for untrained vs. trained on inconsistent metaphors, b = -273.4 54.6 SEM, t = -5.0, p < .0001), with congruency effects with respect to natural language metaphors getting progressively smaller across participant groups.

3.7.5 Comparison to Experiment 1, Metaphor Training for English Speakers

Next we compared performance of Chinese-English bilinguals who were trained on vertical spatio-temporal metaphors in the lab to the performance of English speakers trained on the same metaphor systems in Experiment 1. This allowed us to see whether
the new system of metaphors shaped representations to different extents depending on whether participants were already familiar with a system of metaphors using the vertical axis.

The first model included congruency (always with respect to newly-learned metaphors), interference type, and experiment as fully-crossed fixed effects. There was an overall congruency effect with faster responses for new metaphor-congruent trials than incongruent (\(b = -76.0\ 24.5 \text{ SEM, } t = -3.1, p = .002\)), and that congruency effect did not differ by experiment (\(b = 20.1\ 46.7 \text{ SEM, } t = 0.4, p = .67\)). In other words, newly-learned metaphors did not shape mental representations for Chinese speakers, who have had exposure to vertical spatio-temporal metaphors in natural language, less than it did for English speakers, for whom vertical linguistic metaphors for talking about time are novel.

A second model was identical to the first, but instead of congruency as a fully-crossed fixed effect, it included task condition. This model allowed us to investigate whether an early-up bias manifested to different extents in the two groups, as would be expected if natural language metaphors can shape mental representations even after learning a new representation in language. Indeed, Chinese-English bilinguals who were trained on the new metaphors showed a stronger early-up bias than their English-speaking counterparts (\(b = 107.2\ 47.9 \text{ SEM, } t = 2.2, p = .03\)), suggesting that long-term language influences can affect performance even when people are introduced to a new representation through linguistic metaphors, which simultaneously shapes mental representations.
3.8 **Experiment 4 Discussion**

In this experiment, we taught Chinese-English bilinguals systems of vertical metaphors (in English) for talking about time before completing the nonlinguistic space-time association task used in Experiments 1 and 3 of this dissertation. These participants then demonstrated mental representations that were consistent with their newly-learned metaphors, and these representations were unaffected by verbal (or spatial) interference. The same participants simultaneously demonstrated mental representations consistent with Chinese natural language metaphors, which place earlier events above later ones, and these representations were similarly unaffected by the addition of interference.

These findings suggest that learning a new system of linguistic metaphors for talking about time and long-term exposure to linguistic metaphors for time can create representations that are similar in magnitude and nature (not requiring linguistic working memory resources in the moment).

3.9 **General Discussion**

The experiments presented here explore the effects of long-term and immediate language experiences on mental representations of time. These experiments build on findings from Experiment 1, which suggested that learning a new spatio-temporal system of linguistic metaphors in the lab can create new non-linguistic mental representations.

Experiment 3 was designed in part to examine the external validity of those prior results: Are participants’ natural language-consistent mental representations non-linguistic, as those created in the lab were? To test this, we recruited Mandarin Chinese speakers (who participated in either Mandarin or in English), whose natural language uses vertical metaphors to talk about time (placing earlier events above later ones), and English speakers as a control. We used the same materials as in Experiment 1 to probe
participants implicit space-time associations on the vertical axis and to apply either a verbal (or spatial) dual task while participants made time judgments.

Overall, participants showed an early-up bias, indicated by faster response times when the earlier key was above later than the reverse. This bias was stronger for the Chinese-in-English group than the English group, but did not differ between the two groups of Chinese speakers. This finding demonstrates that participants with natural language experience that includes vertical metaphors think about time in consistent ways to a greater extent than those lacking that same natural language experience.

Further, Chinese speakers’ natural language-consistent representations were unaffected by verbal (or spatial) interference, suggesting that they are non-linguistic in nature, as was the case for representations that arose from new linguistic metaphor systems in Experiment 1 of this dissertation. Although this experiment was necessarily quasi-experimental, since we could not manipulate the variable of natural language experience, it lends weight to the findings of Experiment 1. The current experiment suggests that just as representations shaped through a linguistic training in the lab did not rely on verbal working memory stores to manifest in behavior, neither did those consistent with natural language metaphors.

In Experiment 4, Chinese-English bilinguals learned a system of spatio-temporal metaphors in English that was either congruent or incongruent with their natural language metaphors. They made the same time judgments under the same interference conditions as in the previous experiments. On that task, participants showed both representations consistent with their long-term Chinese experience and with the new metaphors learned in the lab. Those who learned the system consistent with Chinese natural language metaphors showed especially strong early-is-up biases, greater in magnitude than the biases observed in Chinese-English bilinguals who were not trained on any new system in the lab. On the flip side, participants who learned the system inconsistent with natural
language metaphors did not show the early-is-up biases that were observed for Chinese-English bilinguals untrained in the lab, suggesting that the natural language-inconsistent training also impacted the strength of participants’ specific space-time associations.

Of central interest to our research questions, we found that congruency effects, both with respect to Chinese natural language metaphors, and newly-learned metaphors, were unaffected by verbal (or spatial) interference. Thus, effects of immediate and long-term language experiences seem to be nonlinguistic in nature, consistent with the findings from Experiments 1 and 3.

Research showing differences in cognition that are consistent with linguistic differences is necessarily correlational, and therefore is limited in its ability to demonstrate that language shapes thought. Training participants to use new linguistic structures in the lab is one way to address this difficulty, but explicitly teaching people new metaphors in a lab setting has its own drawbacks, since it does not reflect the gradual and long-term exposure that characterize natural language learning.

The first three experiments in this dissertation reconcile the advantages and disadvantages of both studying representations consistent with natural language experience and training people on new metaphor systems in the lab. In Experiment 1 (Ch. 2), we relied solely on training participants to use new metaphors for the experimental control it gave us, which in turn allowed us to make causal inferences about the impact of metaphors on mental representations of time. In Experiment 3 (this chapter), we relied solely on natural language experience since participants were all Chinese-English bilinguals who were not trained in the lab to use any particular metaphors, which allowed us the ecological validity of examining representations without experimental interference.

Until now, it has not been clear how similar newly-learned representations are to those that have developed over a lifetime of language experience. Thus, in Experiment 4, we combined these two methods by relying again on Chinese-English bilinguals, but also
teaching them new systems of metaphors in the lab. This final experiment, in conjunction with the two prior ones, suggested that learning a new way to talk about time in the lab can create mental representations that behave like those acquired through long-term language use, both in magnitude and in their persistence under verbal interference. In doing so, this work opens the door for future investigations of the nature of mental representations we acquire through language.
Chapter 4

Consistencies in space-time and space-number associations under spatial interference

4.1 Abstract

Do space-time associations arising from visuomotor experience differ in nature from those arising from linguistic experiences? The four experiments in this chapter address this question by examining associations on the lateral axis under verbal and visual interference. Together, they examine the generalizability of the prior findings that associations shaped by newly-learned or long-term language experiences are robust to verbal and spatial interference.
4.2 Introduction

The prior experiments in this dissertation have shown that linguistic metaphors – whether learned in the laboratory or through native language experience – can shape behavior on a non-linguistic space-time association task. Further, these associations are not affected by verbal or spatial interference.

Do representations shaped by language differ in nature from those shaped by visuomotor experience? It is possible that space-time associations arising from different sources also differ in their underlying mechanisms; namely, whether verbal or spatial working memory resources are required for them to manifest in behavior. In this case, we might find that associations consistent with visuomotor (reading and writing) experience, which place earlier events to the left of later ones, may not persist under the interference types that language-induced associations did persist under in the prior experiments.

The studies reported here investigate the role of spatial working memory in particular in reasoning about time (Exps 5-8) and number (Exps 6 & 7) by applying a spatial interference task while people make temporal or numerical judgments. Participants made judgments with lateral response keys, such that in the congruent condition the earlier or smaller response was to the left of the later or larger, and in the incongruent condition the earlier or smaller response was to the right.

As in the prior experiments, we consider three possibilities: the application of spatial interference may (1) amplify or (2) reduce the size of congruency effects, demonstrating the involvement of spatial working memory resources for the associations to manifest in behavior. On the other hand, the dual task may (3) not affect the size of the congruency effects, consistent with the long-term account, which suggests that because the associations are at least in part stored in long-term memory, they do not require spatial working memory resources in the moment when the associations are evident in behavior.
The reduction effect in particular has been observed in prior work on space-number associations [67, 122]. For example, systematic space-number associations were not observed under spatial interference. Without interference, Belgian university students were faster to make magnitude judgments about single-digit numbers (i.e., larger or smaller than 5?) when the response buttons were congruent (rather than incongruent) with a left-right number line: when the smaller response was to the left of the larger response, demonstrating the SNARC effect, which had been documented previously [34]. Under spatial interference (using the Corsi task), however, the SNARC effect was eliminated [67, 122]. As mentioned previously, the authors explain the elimination of the SNARC effect under spatial interference by suggesting that people temporarily represent spatial information associated with different numbers in spatial working memory [67]. These temporarily-maintained spatial properties give rise to the SNARC effect, but are not responsible for people’s ability to make numerical judgments. Thus, their ability to respond is uninhibited, but they are no longer faster for congruent (smaller numbers to the left of larger ones) than incongruent judgments. Because of the similarities between mental space-time and space-number associations, these studies provide an additional reason to predict that lateral space-time associations might be disrupted by spatial interference.

However, it is also possible, if spatial working memory resources are necessarily invoked in the moment when space-time associations are observed, that spatial interference could amplify congruency effects. In this case, spatial interference could differentially affect people's ability to make congruent and incongruent judgments. Incongruent judgments may become especially difficult as people work to maintain the key mapping (which conflicts with their canonical representation stored in long-term memory) in working memory along with the additional spatial information needed for the interference portion of the trial. As a result, incongruent reaction times might be
especially slow, in turn increasing the difference between congruent and incongruent trials, or, in other words, leading to a greater congruency effect under spatial interference. This outcome could suggest that the introduction of new spatial information (i.e., the incongruent key mapping instructions) taxes working memory. In this paper, we refer to this possibility as the amplification effect. In short, both the reduction and amplification of congruency effects would demonstrate that spatial working memory is necessarily recruited when reasoning about time.

On the other hand, a body of research supports a third possibility, which we refer to as the long-term memory account. Under this account, the manifestation of space-time associations does not require short-term memory stores in the moment, so a spatial working memory dual task would not interfere with lateral space-time associations. For example, in prior work English speakers who learned a system of linguistic metaphors not present in their native language (earlier events were said to be either above or below later ones) demonstrated new metaphor-congruent ways of thinking about time that were unaffected by verbal or spatial interference (Experiment 1 [65]). The newly-learned representations didn’t require either of these working memory stores in the moment to manifest, since the size of the congruency effects was unchanged under the different interference types. If lateral space-time associations are similar in nature to those on the vertical axis acquired through linguistic metaphor, we may not find an interference effect from applying a spatial dual task when people make time judgments with lateral responses.
4.3 Experiment 5: Space-Time Associations from Visuo-motor Experiences

Are mental representations of time that result from reading and writing experience robust to verbal and spatial interference, as those arising from linguistic metaphor in Experiment 1 were? To address this question, we adapted the methods used in Experiment 1 to test American college students’ left-right mental timelines.

4.4 Experiment 5 Methods

4.4.1 Participants & Trial Inclusion

A total of 102 UC San Diego undergraduates participated for course credit. They received a link to participate through the Psychology participant recruitment site. Participants were excluded based on the same inclusion criteria as in Experiment 1. After exclusions, the three conditions included 30 participants each, for a total of 90 participants.

4.4.2 Materials & Procedure

The materials and procedure were identical to those of Experiment 1 with two exceptions. Participants did not learn a new way of talking about time, and the response keys were arranged on the left-right axis.

As in Experiment 1, participants in the interference conditions completed the same calibration block at the beginning of the experiment, and everyone completed two blocks of the time judgment task with opposite key mapping instructions. The time judgment task included either spatial or verbal interference for participants in those
conditions. There were again 10 practice trials and 56 experimental trials in each block.

### 4.4.3 Analyses

We again fit the data with mixed effects models. The full mixed effects model included congruency (with the canonical left-right timeline), interference condition, and block order (whether the congruent block was the first or second time judgment block) as fully crossed fixed effects, and subjects and items as random effects. As in Experiment 1, we compared reduced models to the full model to investigate main effects and interactions.

### 4.5 Experiment 5 Results

#### 4.5.1 Judgment Reaction Times

American college students showed robust left-right mental timelines that were not changed by either verbal or spatial interference.

Participants were faster when the response keys were arranged with *earlier* on the left (1578 ms) than with *earlier* on the right (1968 ms), as confirmed by a main effect of congruency, $X^2(1)=51.8$, $p < .00001$. Table 4.1 shows mean reaction times for each group in each time judgment block.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Verbal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT (ms)</strong></td>
<td>congruent</td>
<td>1430 (450)</td>
<td>1770 (415)</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>1805 (631)</td>
<td>2129 (483)</td>
</tr>
<tr>
<td><strong>Judgment Accuracy (%)</strong></td>
<td>congruent</td>
<td>93.3</td>
<td>92.8</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>91.8</td>
<td>90.3</td>
</tr>
</tbody>
</table>

*Table 4.1: Left-Right Reaction Times (SD) and Time Judgment Accuracies by Interference Type*
As in Experiment 1, each participant’s congruency effect was calculated by subtracting their mean reaction time on congruent (earlier = left; later = right) trials from their mean reaction time on incongruent trials (shown in Figure 4.1). Congruency effects did not differ among the interference types, as there was no interaction between interference type and congruency, \(X^2(2) = 0.47, p = .79\). Pairwise comparisons of the congruency effect between all possible interference types revealed the same null effect, (all ps > .50). In sum, manifestations of people’s left-right mental timelines were unaffected by the interference conditions they experienced while making time judgments.

![Figure 4.1: Results of Experiment 5: Congruency effect for representations on the left-right axis (ms). Congruency effects were calculated for each participant by subtracting their mean reaction time for congruent trials (earlier response key is to the left of the later key) from their mean reaction time for incongruent trials.](image)

### 4.5.2 Accuracies

People were overall more accurate on congruent time judgment trials (92.4%) than on incongruent trials (89.6%); \(F(1,87) = 14.93, p = .0002\). This suggests that the congruency results observed in reaction times were not due to a speed-accuracy tradeoff,
as participants were both faster and more accurate on congruent trials.

There was a main effect of interference type on time judgment accuracies (F(2,87) = 4.73, p = .01). Pairwise tests revealed that participants in the spatial interference group were less accurate on the time judgment task (88.8%) than the participants in the verbal interference (91.6%; t(101.4)=2.36, p = .02) and no interference (92.6%; t(110)=3.08, p = .003) groups. There was no difference in accuracy between the verbal and no interference groups (t(115.4)=1.04, p = .30).

The number of distractor items assigned to the participants for the time judgment task based on calibration was the same for both the verbal (M=4.47, SD = 1.14) and spatial (M=4.90, SD = 1.42) interference groups (p = .20).

Accuracies on the interference task were higher for congruent trials (62.7%) than incongruent trials (58.3%; F(1,58) = 6.05, p = .02). They were also higher for participants who experienced verbal interference (69.8%) than those who experienced spatial interference (51.2%; F(1,58) = 14.2, p = .0003). Accuracies for each interference group on both congruent and incongruent trials are shown in Table 4.1.

4.5.3 Comparing Behavior Shaped by Learning New Metaphors (Exp 1) to Behavior Shaped by Visuomotor Experience (Exp 5)

We combined the data from Experiments 1 and 5 to investigate differences in the congruency effects that arise from linguistic metaphor (Exp 1) and from visuomotor experience (Exp 5). Although participants demonstrated a larger congruency effect on the left-right than on the vertical axis (X^2(1) = 75.1, p < .0001), the interference types did not differently influence the congruency effect between the two studies (X^2(2) = .87, p = .65). In other words, there was no difference in the robustness of peoples spatiotemporal representations to the interference types between the group whose representations arose from new metaphors and the group whose representations arose from visuomotor
experience.

Our studies included another kind of interference intrinsic to the experiment design. On incongruent blocks, participants were asked to use key mappings that contradicted their representations (whether newly learned or long established). This instruction can itself be viewed as a kind of training - creating a new association between space and time. This allowed us the opportunity to ask whether this new instruction also had an effect on people’s representations, and if so whether that effect was different for those whose representations resulted from learning new vertical metaphors than those whose representations resulted from a lifetime of visuomotor experience. To examine this, we computed the size of the congruency effect across trials. That is, as participants completed more trials with either the congruent or incongruent instruction, did the congruency effects decrease, increase, or stay the same? If the response instruction is indeed providing a kind of counter-training, we might expect that the congruency effect should be stronger in earlier trials and weaker in later trials, decreasing as the number of counter-training trials increases. This is indeed what we found for both experiments, as shown in Figure 4.2. There was a significant negative correlation between trial number and the congruency effect both in Experiment 1 ($r(54) = -0.32$, $p = .02$), and in Experiment 5 ($r(54) = -0.35$, $p = .009$). These findings suggest that the instructions for a block of time judgments act as a separate kind of training, strengthening the instruction-consistent representation of time over the course of a block.

It is striking that even representations that have been established over many years of experience (here, the left-right representations in Experiment 5), can be changed or partially overridden by a new mapping introduced in the lab (here, the arrangement of response keys). These findings, along with the results of Experiment 1, suggest that our representations of time are malleable and dynamic. People appear willing and able to learn new psychologically real spatializations for time through a variety of means.
Figure 4.2: Congruency effect (in ms) plotted over the course of time judgment blocks for Experiments 1 & 5. Congruency effect (faster responses for congruent blocks than incongruent blocks, whether with respect to newly-learned metaphors or visuomotor experience) over the course of time judgment blocks.

(whether through verbal metaphor or through a pattern of motor movements to response keys).

4.6 Experiment 5 Discussion

In this experiment, space-time congruency effects on the lateral axis persisted under spatial (and verbal) interference. This finding, consistent with the long-term memory account, suggests that people’s specific space-time associations, conceptualizing earlier events to the left of later ones, do not rely on the spatial attentional resources that a traditional spatial interference task requires.

Since space-time associations arising from linguistic metaphor have also persisted under verbal and spatial interference in prior work (Experiment 1 [65]), the findings
from this experiment suggest that mental representations arising from different origins (linguistic vs. reading and writing experience) do not necessarily differ in nature.

Since space-time congruency effects were unaffected by interference, we next asked how generalizable this finding is to other lateral representations; namely, space-number associations. An initial comparison of our results with prior work showing that space-number congruency effects can be disrupted by spatial interference [67, 122] suggests that mental representations of time and number may differ in their persistence under spatial interference. However, Experiment 5 did not directly compare temporal and numerical associations, so there are many potential explanations for the apparent discrepancy between the persistence of space-time associations in our work (suggesting the long-term memory account) and the susceptibility of space-number associations to spatial interference in prior work (demonstrating a reduction effect [67, 122]). Thus, in Experiments 6 & 7 we subject space-number associations to similar interference tasks to those used in Experiment 5 to ask: do space-number associations show the same persistence under interference that space-time associations did in the previous experiment?

4.7 Experiments 6 & 7: Mental Time- vs. Number Lines

We look to mental representations of number because they share spatial features like orientation and direction with representations of time.\(^1\) We consider the same potential outcomes (the amplification, reduction, and long-term memory accounts) for space-number associations. Evidence for the spatial nature of thought about number comes in part from research showing a common neural coding for spatial and numeric associations.

\(^{1}\)These experiments are reported in a manuscript in preparation: Hendricks, R.K., Walker, E.J., Bergen, B.K., & Boroditsky, L. Consistencies in space-time and space-number associations under spatial interference.
tasks. For example, Gobel et al. [54] found that when repetitive Transcranial Magnetic Stimulation (rTMS) was applied to the right posterior parietal lobe (a brain area implicated in spatial cognitive tasks), people’s mental representation of number was altered. Specifically, their number bisection estimations (i.e., what number is between 723 and 781?) shifted to the right. Other work has also suggested that common brain areas (specifically, the angular gyrus) are involved in the interaction between space-number associations and spatial attention [21].

Similarly, thinking about numbers has affected the way people allocate spatial attention in a visual detection task [40]. In this work, people saw a number and immediately after had to detect a target stimulus on a screen. After small numbers, they were faster to detect a stimulus on the left side of space, and vice versa, suggesting that simply considering a number biased people’s attention to the location in physical space that corresponds to the number’s location on a mental number line.

Although thought about time and number both demonstrate spatial features, there are also reasons to expect we might not observe behavior consistent with the long-term memory account for the domain of number as we did for time in Experiment 5. For example, prior research also points to differences in these two types of representations, calling into question whether the role of spatial working memory in the manifestation of space-time and space-number associations will be consistent. For example, Bottini et al. [16] suggested that time and number are mentally represented in different types of space. By examining both types of representations in sighted and blind participants, they found that lateral space-time associations develop even without vision, and that blind and sighted people all base this representation on an external frame of reference. This finding contrasts with prior work showing that early-blind people’s lateral space-number associations are anchored on their bodies, while late-blind and seeing participants’ representations are anchored in external space [31]. In other words, developing an
externally-anchored mental representation of time seems to happen in the absence of visual information, while an externally-anchored mental representation of number seems to rely on visual experience.

Another reason to predict that our findings may not be consistent with the long-term memory account for space-number associations under spatial interference comes from work that used similar methods to the ones applied here. Specifically, van Dijck et al. [122] found that when participants made parity judgments (is this number even or odd?), space-number congruency effects were unaffected by a spatial dual task. However, when participants made magnitude judgments (is the number larger or smaller than 5?) while completing the same same spatial working memory task, the researchers observed a reduction in the size of the congruency effects, compared to when participants made the same judgments while engaging in a verbal working memory task. This experiment, along with others that show an overlap between thinking about number and engaging spatial cognitive resources, suggests that the behavior consistent with the long-term memory account we observed for making time judgments under spatial interference in the prior experiments may not be found when participants make numerical magnitude judgments—instead, time and number representations may differ in their reliance on spatial working memory resources.

To investigate the generalizability of our finding from Experiment 5 that mental representations of time persist under spatial interference, Experiments 6 & 7 conceptually replicated and extended the prior study. Participants again made judgments under either verbal or spatial interference, but while some made time judgments, others made numerical magnitude judgments.

Do space-time and space-number associations differ in their nature—specifically, in the extent to which they manifest under spatial interference? If so, in Experiments 6 & 7 we would expect specific space-time associations to remain unaffected under
spatial interference, while space-number associations may be reduced (or amplified) under spatial interference, as in prior work [67, 122].

4.8 Experiments 6 & 7 Methods

The methods for Experiments 6 and 7 were identical with one exception. In Experiment 6, participants completed two time or number judgments between encoding and recalling the interference distractors, while in Experiment 7, they completed sixteen. We included this variable in part to more closely align our methods to those of van Dijck et al. [122]. In addition we aimed to increase the difficulty of the interference task by including sixteen judgments between encoding and recall, in case maintaining spatial information in mind while only completing two judgments was not taxing enough for us to observe interference effects. We differentiate this work into two experiments because they were not conducted in tandem, so participants were not randomly assigned to make either 2 or 16 judgments between encoding and recall. After exclusions and counterbalancing we arrived at a different number of participants for Experiments 6 and 7.

Since Experiments 6 and 7 differed only in the number of judgments between encoding and recall, we combined the data and included the number of trials between encoding and recall as a between-subjects factor. Methods and Results are therefore reported together.

4.8.1 Participants & Trial Inclusion Criteria

Two hundred fifty-eight undergraduates recruited through the UCSD Psychology Department participant pool received a link to participate on their own computers for course credit.
Twenty-six participants were excluded for responding with an accuracy of 50% or lower on at least one time or number judgment block. After exclusions, there were 96 participants in Exp. 6 and 136 participants in Exp. 7. Only correct judgments (94.5%) and trials with reaction times greater than 250 ms and within 3 standard deviations of each participant’s cell mean (97.0% of correct responses) were analyzed.

4.8.2 Materials and Procedure

The experiment was programmed using jsPsych [33]. The study included an interference calibration block (detailed in Interference Calibration), followed by two time or number judgment blocks with interference. Each participant was randomly assigned to make time or number judgments under verbal or spatial interference, for a total of four between-subject conditions. Whether the first judgment block was congruent or incongruent was counterbalanced across all conditions.

The first and second time/number judgment blocks had opposite assignments of response keys: in one block, the earlier/smaller key was to the left of later/larger and in the other, the key assignments were reversed. Block order was counterbalanced.

Interference Calibration

The calibration procedure was identical to the one used by Frank et al [42] and in the prior experiments in this dissertation. The MATLAB code to run the interference tasks was downloaded from https://github.com/langcog/numint and modified to run using jsPsych. For verbal interference, participants saw a string of letters, then completed a visual search task, and finally typed the letters. For spatial interference, letters were replaced by a 4x4 grid of white squares, and individual squares flashed blue one at a time. Participants started by remembering 2 letters or boxes, and difficulty was calibrated in a 2-up, 1-down staircase.
4.8.3 Time Judgments

Participants pressed a central key on their keyboard (‘G’) to begin, and the first image appeared (e.g., Julia Roberts in her 20s). After 2000 ms, this image was replaced by the second image (e.g., either a younger or older Julia). Participants determined whether the second image showed an earlier or later time point than the first and pressed a key corresponding to their decision. The earlier key was to the left of the starting key in one block (‘D’), and to the right in the other (‘J’).

Number Judgments

For the numerical magnitude judgments, participants always saw a ‘5’ for 2000 ms, followed by a second digit. They indicated whether the second digit was smaller or larger than 5, pressing the same response keys as in the time judgment condition.

Dual Task

Interference levels from calibration were used so that the tasks were tuned for individual ability. Participants completed 9 practice judgments, 5 without interference and 4 with interference. For the main task, interference presentation was followed by 2 (Experiment 6) or 16 (Experiment 7) judgments (time or number) before participants recalled the interference items.

In Experiment 6, there were 56 judgments (28 interference trials) per block and in Experiment 7, there were 48 judgments (3 interference trials). Whether the correct response was earlier or later was randomized.
4.8.4 Analyses

We analyzed the data with mixed effects models. We used the maximal random effect structure that would converge. Interference type (verbal, spatial), congruency (congruent, incongruent), domain (time, number) and number of trials between encoding and recall (2, 16) were included as fixed effects. Participants and items were included as random intercepts.

4.9 Experiments 6 & 7 Results

4.9.1 Judgment Reaction Times

Overall, people were faster to make congruent (1,044 ms) than incongruent (1,205 ms) judgments ($b = 81.5$, $SE=4.07$, $t=20.0$, $p < .0001$), replicating extensive work showing that American English speakers exhibit behavior consistent with left-to-right space-time [120] and space-number associations [34]. Table 4.2 shows reaction times for all groups and conditions and Figure 4.3 shows congruency effects.

Participants who made number judgments (839 ms) were faster than those who made time judgments (1410 ms; $b = -312.5$, $SE=87.6$, $t=-3.57$, $p = .0006$), likely because digits are less visually complex than the images used in the time judgment task, and because every digit had to be repeated for the number judgments, while the pictures used for time judgments were not repeated. The congruency effect was greater for time (245 ms) than number (77 ms) judgments ($b = -41.9$, $SE=4.07$, $t=-10.3$, $p < .0001$), likely because reaction times were overall greater for time than number judgments.
Table 4.2: Reaction times for Experiments 6 and 7 in ms (SD) by number of trials between interference encoding and recall and interference type. All reaction times are in ms.

<table>
<thead>
<tr>
<th></th>
<th>2 trials</th>
<th>16 trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>spatial</td>
<td>verbal</td>
</tr>
<tr>
<td>number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>congruent</td>
<td>806 (230)</td>
<td>875 (245)</td>
</tr>
<tr>
<td>incongruent</td>
<td>892 (262)</td>
<td>969 (255)</td>
</tr>
<tr>
<td>time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>congruent</td>
<td>1512 (366)</td>
<td>1630 (548)</td>
</tr>
<tr>
<td>incongruent</td>
<td>1757 (486)</td>
<td>1859 (526)</td>
</tr>
</tbody>
</table>

Figure 4.3: Results of Experiments 6 & 7: Congruency effects by interference type. Number judgments are in the left panel and time judgments are in the right.

4.9.2 Judgment Accuracies

Participants were overall more accurate for congruent (95.3%) than incongruent (93.1%) judgments ($b = -.198$, $SE = .036$, $z = -5.521$, $p < .001$). Table 4.3 shows time and number judgment accuracies.
<table>
<thead>
<tr>
<th>number</th>
<th>spatial</th>
<th>verbal</th>
<th>spatial</th>
<th>verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>congruent</td>
<td>98.3</td>
<td>96.1</td>
<td>97.4</td>
<td>97.5</td>
</tr>
<tr>
<td>incongruent</td>
<td>96.1</td>
<td>98.3</td>
<td>97.1</td>
<td>97.5</td>
</tr>
<tr>
<td>time</td>
<td>congruent</td>
<td>92.9</td>
<td>91.5</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>91.9</td>
<td>86.1</td>
<td>93.1</td>
</tr>
</tbody>
</table>

### 4.9.3 Interference Accuracies

Participants were overall more accurate at remembering the interference distractors when they accompanied congruent (68.0%) as opposed to incongruent (66.3%) judgments ($b = -0.11$, $SE=0.04$, $z = -2.37$, $p = .02$). This congruency effect was greater under spatial interference (3.6% more accurate for congruent than incongruent trials) than verbal (0.3% difference; $b =0.09$, $SE=0.05$, $z = 1.98$, $p = .05$). In other words, participants had greater difficulty with the spatial dual task (compared to the verbal) when they also had to maintain incongruent information about the spatial layout of the response keys in mind. This finding supports the amplification account, in which spatial working memory plays a role in the manifestation of space-time associations specifically when the task requires participants to maintain information in mind that conflicts with their canonical space-time associations (making incongruent judgments).

The interference accuracy congruency effect was also greater when people made 16 judgments between encoding and recall (7.8% difference) than 2 (0.7% difference; $b = -0.09$, $SE=0.05$, $z=-1.95$, $p = .05$). Thus, incongruent trials were especially challenging with an increased delay between encoding and recall. There was also an interaction between interference type and number of trials between encoding and recall ($b = 0.21$, $SE=0.88$, $z=2.33$, $p = .02$). Follow-up tests showed that for verbal interference, participants were more accurate when there were 16 judgments than 2 (7.6% difference), but
for spatial interference the reverse pattern was observed: people were more accurate
when there were 2 judgments than 16 (3.0% difference). Again, we interpret this finding
as a potential indication of an amplification effect – specifically, that spatial working
memory involvement in space-time associations may be evident when the task requires
participants to maintain extra spatial information in mind.

Participants were more accurate for recalling verbal (75.0%) than spatial (59.2%)
interference distractors (b = 0.53, SE=0.08, z=6.85, p < .0001), mirroring the results of
prior work by Frank et al. [42]. Participants were also more accurate for interference
accompanying number (72.3%) than time judgments (62.0%; b = 0.17, SE=0.08, z=2.09,
p = .04), which is consistent with the finding that people were faster to make number
than time judgments, likely because the number stimuli were less visually complex than
the time stimuli, and because digits were necessarily repeated multiple times throughout
each block, while images for the time judgments were not. Interference accuracies are
shown in Table 4.4.

Table 4.4: Interference accuracies for Exps 6 and 7 (%) by interference type and number
of trials between encoding and recalling interference distractors.

<table>
<thead>
<tr>
<th></th>
<th>2 trials</th>
<th>16 trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>spatial</td>
<td>verbal</td>
</tr>
<tr>
<td>number</td>
<td>67.5</td>
<td>77.2</td>
</tr>
<tr>
<td>time</td>
<td>51.8</td>
<td>70.8</td>
</tr>
</tbody>
</table>

4.10 Experiments 6 & 7 Discussion

In these experiments, we asked whether space-number associations persist under
the same interference conditions (consistent with the long-term memory account) as
space-time associations did in the prior experiments (Exps 1, 3, 4, & 5). Participants
made number (larger or smaller?) or time (earlier or later?) judgments while response
keys were either congruent (smaller/earlier to the left of larger/later) or incongruent with canonical space-time and space-number associations. Each person made these judgments under either verbal or spatial interference. Finally, some participants made 2 judgments between encoding and recalling the interference distractors while others made 16.

We found reaction time congruency effects for both time and number; that is, people were faster to make judgments when the response keys were congruent with the canonical left-right space-time and space-number associations than incongruent. Further, congruency effects for both time and number judgments were unaffected by spatial (or verbal) interference, suggesting one way in which these two types of mental representations do not differ: neither space-time nor space-number congruency effects require the spatial attentional working memory resources that are recruited by a traditional spatial working memory task in order to manifest. The behavior indicative of the long-term memory account we observed here further mirrors findings from earlier work on mental representations of time arising from linguistic metaphor [65], suggesting that space-time associations, whether shaped by reading and writing experience or by language, are unaffected by a spatial working memory dual task.

Reaction time congruency effects were also unaffected by the number of judgments participants made between encoding and recalling the interference distractors, suggesting that space-time and space-number associations dont manifest to a different extent whether people make many subsequent judgments or just a couple during a single series. Further, the spatial dual task did not differentially affect these associations when it had to be maintained in working memory for many subsequent trials or a couple. As a result, we can rule out the potential explanation that the lack of an effect of spatial interference on congruency effects we observed in the prior experiments was because the spatial task wasn’t sufficiently challenging, such that people did not have to maintain the spatial information in mind for long.
However, judgment reaction times for participants who completed 2 intervening judgments were overall slower than for those who completed 16. This may be a counterintuitive finding, since presumably completing 16 judgments between encoding and recalling interference stimuli is more challenging than only making 2 judgments. We suspect that it is precisely because it was more challenging to maintain interference distractors in memory for 16 judgments that participants were faster to make the judgments in this condition – they had greater motivation to complete the judgments so they could offload the distractors they were maintaining in memory. Another potential explanation for the difference in judgment reaction times between the two interference delay conditions is that those who made 16 judgments between encoding and recalling the interference stimuli had fewer sets of interference distractors to remember overall (3 sets per block) than those who made only two judgments between encoding and recalling the stimuli (24 sets per block). This difference in the number of interference trials experienced arises from the fact that the two groups made the same number of judgments throughout the experiment; varying the number of judgments between encoding and recall made it impossible to have participants all complete the same number of interference trials. Thus, those who had an interference delay of only 2 judgments may have been more susceptible to fatigue than participants who had an interference delay of 16 judgments, contributing to increased judgment reaction times in the 2-trials-between condition than in the 16-trials-between condition.

Although reaction time congruency effects were not differentially affected by the two dual tasks (consistent with the long-term memory account), accuracies on the interference task hint at the possibility of differential involvement for spatial and verbal working memory for space-time associations to manifest in behavior. Specifically, we found that under spatial interference, compared to under verbal interference, participants were less accurate at recalling the interfering information for incongruent trials than for congruent;
we saw a greater congruency effect for interference accuracies under spatial than verbal interference. In other words, we observed an amplification effect on interference memory accuracy when spatial information was maintained in working memory while participants made temporal judgments. It seems possible, then, that participants may draw on spatial working memory resources to some extent when canonical space-time and space-number associations manifest in behavior, though much remains unknown about the manner in which spatial working memory is involved in this process (if it is). Since we did not observe an impact of spatial interference on reaction time congruency effects in this same way, the extent to which spatial working memory may be required for determining the relative order of events remains an open question for future work.

4.11 Experiment 8: Mental Timeline under Horizontal vs. Vertical Interference

In Experiments 5-7, space-time associations between earlier events and the left side of space and later events and the right were unaffected by spatial interference. Similarly, vertical space-time associations shaped through metaphors in language in Experiment 1, as well as those consistent with patterns in language in Experiments 3 and 4, were not disrupted by spatial interference. We interpreted the prior results under spatial interference as suggesting people do not need to invoke spatial working memory resources in order to demonstrate space-time associations in behavior; instead, space-time associations are likely stored, at least to some extent, in long-term memory.

However, it is also possible that the spatial interference paradigm we used (a 4x4 grid of boxes) didnt tax working memory for the most relevant aspect of space. Mental timelines may rely on a more specific subset of spatial working memory, focused on the relevant axis. In the case of left-right mental timelines, then, a dual task that...
requires people to maintain specifically lateral spatial information in working memory may affect space-time associations more than one that does not. Thus, this experiment tested whether space-time associations on the lateral axis would be differentially affected by a relevant axis-specific and an irrelevant axis-specific dual task. In other words, the spatial information that participants maintained in working memory was either oriented laterally (the same axis as the response keys they used to make time judgments) or vertically.

This experiment was identical to Exp 5, in which participants made temporal judgments in response to pictures by pressing laterally-oriented response buttons. However, instead of completing either spatial (4x4 square) or verbal interference while making the time judgments, all participants completed either an 8x1 (vertical) or 1x8 (horizontal) interference task. If invoking left-right space-time associations to make time judgments relies on working memory along the left-right axis, the congruency effect should be affected by the lateral spatial interference dual task in ways that it is not affected by the vertical dual task. We may observe that lateral space-time associations (congruency effects) are reduced when people make judgments accompanying lateral spatial information if the working memory task interferes with their ability to construct a mental timeline; without a mental timeline, there may be no difference in reaction times when the trial is congruent versus when it is incongruent with canonical timelines. Conversely, if lateral-specific working memory is recruited when people invoke a lateral mental timeline, the space-time associations could also be amplified, for example if maintaining lateral spatial information in mind makes incongruent time judgment trials especially taxing compared to congruent ones.
4.12 Experiment 8 Methods

4.12.1 Participants & Trial Inclusion

One hundred thirty-six undergraduates at UCSD participated for course credit. Twenty-five were excluded for interference accuracies on at least one block of time judgments that were below 25%. Seven others were excluded for time judgment accuracies on at least one block that were lower than 3 standard deviations from the group mean accuracy. After exclusions, we counterbalanced so that for each axis, an equal number of participants completed each block order. This left 50 participants in the horizontal interference group and 46 in the vertical interference group.

Only trials for which interference memory was correct (68.7% of trials), time judgment was correct (91.2% of correct interference trials), and reaction times were between 250 and 5,000 ms and within 3 standard deviations of each participant’s cell mean (96.5% of all accurate trials) were included.

4.12.2 Materials & Procedure

The materials and procedure were identical to those in Exp 5 except that instead of a 4x4 grid of boxes, the spatial interference was either vertical (8x1) or horizontal (1x8). The boxes were also slightly smaller than in the prior experiment so that 8 vertically-oriented boxes could be displayed on the screen in a single frame.

As in the prior experiments, we calibrated interference so the task was as comparably difficult as possible for all participants. As in the prior experiments, this took the form of a 60-trial two-up, one-down staircase. Following the interference calibration, on time judgment trials, participants saw a sequence of two images and indicated as quickly as possible whether the second showed an earlier or later time point than the first. Time judgments were preceded by a presentation of the spatial interference (either
oriented vertically or horizontally) and followed by a blank grid for participants to recall the distractors.

4.12.3 Analyses

We conducted generalized linear mixed effects models using the lme4 package in R [7]. Interference axis (vertical, horizontal) and congruency (congruent, incongruent) were included as contrast-coded fully-crossed fixed effects. Number of distractors was also included as a fixed effect. Participants and items were included as random intercepts. For reaction times, we included random slopes by subject for congruency. For time judgment and interference accuracies, we included random slopes by subject for congruency.

4.13 Experiment 8 Results

4.13.1 Interference Distractors

Based on the calibration at the beginning of the experiment, participants in the horizontal interference group did the task with more distractors (M=4.1) than those in the vertical interference group (M=3.1; t(87.2)=2.8, p=.005). For this reason, we included the number of distractors as a fixed effect in all analyses. We also completed separate analyses with a sample that was matched for number of distractors between the two groups. To create this sample, we eliminated 8 participants from each group so that the number of distractors that the remaining participants saw was the same (M = 3.55). The pattern of results did not differ without these 16 participants.
4.13.2 Judgment Reaction Times

Participants were faster to make time judgments when the response keys were congruent with a left-to-right timeline (1618 ms) than incongruent (1830 ms; b =105.2, SE=17.3, p< .0001). There was no difference in reaction times under vertical (1650 ms) and under horizontal interference (1792 ms; b =26.8, SE=25.3, p=.29). The congruency effect also did not differ under vertical and horizontal interference (there was no interaction between congruency and interference axis; b =8.5, SE=15.4, p=.58). Reaction times are shown in Figure 4.4.

Figure 4.4: Results of Experiments 8: Reaction times by axis of spatial interference. All RTs are in ms.
4.13.3 Judgment Accuracies

Time judgment accuracy was not different for congruent (91.4%) and incongruent trials (89.8%; b =.07, SE=.15, p=.62), nor was it different for vertical (90.0%) or horizontal (91.6%) interference (b =.14, SE=.17, p=.41). There was no interaction between congruency and interference axis (b =.17, SE=.21, p=.40).

4.13.4 Interference Accuracies

Interference memory was also not different when it accompanied congruent (68.9%) or incongruent (68.4%) time judgments (b =.08, SE=.10, p=.39). Participants were marginally better at remembering horizontal (70.6%) than vertical (66.6%) interference information (b =.28, SE=.16, p=.08). There was no interaction between congruency and axis of interference (b =.09, SE=.14, p=.50), so the two interference groups did not have different interference accuracy congruency effects.

4.14 Experiment 8 Discussion

When we invoke a left-right mental timeline to reason about sequences of events, do we use cognitive resources devoted to left-to-right spatial working memory? In earlier experiments (5-7), lateral space-time associations were unaffected when participants engaged in a spatial dual task while making time judgments, suggesting that spatial working memory resources do not need to be recruited in the moment for space-time associations to manifest in behavior. The current experiment was designed to test one possible explanation for that previous result – specifically, that general spatial working memory may not be necessary for invoking space-time associations, but that axis-specific spatial working memory may be.

To test for this possibility, the current experiment was identical to Experiment
5, in which participants made earlier/later judgments to picture stimuli by pressing laterally-oriented response keys. All participants experienced spatial interference: half maintained an 8x1 (vertical) grid in memory while making time judgments, while the other half maintained a 1x8 (horizontal) grid in memory. If axis-specific visuospatial mental resources must be invoked so people can construct a spatial representation for the purposes of completing a spatial task, we would expect the horizontal dual task to affect this process in some way – compared to vertical spatial interference, the horizontal interference may alter the size of the congruency effect (amplifying or reducing it), slow people down while making the time judgments, or impair their accuracy for either making time judgments or remembering the interfering information.

In the current experiment, we saw none of these potential disruptions when participants made time judgments under horizontal, as compared to vertical, spatial interference. These results suggest that the reason that space-time congruency effects on the vertical (Exps 1, 3, 4) or horizontal (Exps 5-7) axes in prior experiments were not disrupted by spatial interference was not because we used a spatial dual task focused on an irrelevant or unspecific aspect of space.

These results also point to a potentially important distinction – there’s a difference between the cognitive processes people must engage when doing a task and those they typically engage. Dual task paradigms as used in this experiment should affect performance when the task requires a cognitive process that people must engage for doing some other task. However, dual task paradigms will not necessarily affect performance when the resources required by the dual task are ones that people typically engage. If certain cognitive resources that people would normally use for a particular task are recruited by the dual task, but people can use different resources for the primary task at hand, they are likely to do so. In this case, there may not be performance effects that manifest in a behavioral task.
For this reason, behavioral tasks may be limited for addressing the functional role of space in thinking about time. Future work might look toward neuroimaging methods as a more effective way of testing hypotheses about the cognitive processes involved in thinking about time and the extent to which they involve those for thinking about space.

4.15 General Discussion

The results of four studies suggest that space-time and space-number associations that arise from visuomotor (reading and writing) experience persist under verbal and spatial interference. This was true when the spatial interference took the form of a 4x4 grid of boxes (Exps 5-7, as in Exps 1, 3, & 4) or a horizontal or vertical array (Exp 8). This finding suggests that even when spatial working memory resources are recruited to attend to distractors on the same axis as the relevant space-time association, spatial interference does not affect the size of the associations.

Since earlier work has shown that space-time associations shaped by language experience are also unchanged under interference (Experiments 1, 3, & 4), the pattern of behavior consistent with the long-term memory account we observed in both experiments presented here seems to be a feature of two common spatial mental representations (time and number), even when they are shaped by different sources (i.e., linguistic metaphors and cultural [visuomotor] experiences) and manifest on different axes (vertical and lateral). Across these domains and origins, the existing data suggest that canonical space-time and space-number associations do not require spatial working memory in the moment to manifest.

Our investigation of the persistence of space-time and space-number associations under spatial interference addresses one particular sub-question of a broader cognitive science issue: What mechanisms underlie the spatial properties of our mental repre-
sentations of time and number? However, the findings can inform our understanding of mental representations beyond those for time and number. This work is a first step in a larger enterprise to more specifically articulate the mechanisms that underlie our representations of abstract concepts, including domains like affection, affect, and power. Rather than solely documenting such associations, the experiments described here ask, in more detail, how these associations are stored and implemented in reasoning. These studies pave the way for future work that continues to improve our understanding of our conceptual architecture, including the nature and mechanisms of pervasive associations in development and use.

Chapter 4, in part, is a reprint of the material as it appears in Hendricks, R. K. & Boroditsky, L. (2017). New Space-Time Metaphors Foster New Nonlinguistic Representations, Topics in Cognitive Science. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Hendricks, R. K., Walker, E. J., Bergen, B. K. & Boroditsky, L. (2017). Consistencies in space-time and space-number associations under spatial interference. Acta Psychologica. The dissertation author was the primary investigator and author of this paper.
Chapter 5

Interim Meta-Analyses

5.1 Abstract

The previous three chapters have used overlapping methods to address a suite of related questions: Can metaphors in language create non-linguistic mental representations? Are representations arising from newly-learned metaphors similar in nature to those consistent with long-term natural language use or others driven by cultural practices like reading and writing? Do spatio-temporal and spatio-numeric representations differ in nature? Because these studies used variations on a single set of methods, meta-analyses can shed light on overarching patterns in these mental representations, such as how the representations manifest on different time scales, the role of an individual’s language background on such representations, and what features of the stimuli influence their manifestation.
### 5.2 Methods

Data for Experiments 1 and 3-7 were combined for analyses. For each experiment, this includes only the cleaned data (fully correct trials with reaction times within 3 standard deviations of each participant’s cell mean) that was analyzed in a prior chapter. Figure 5.1 demonstrates the experiments included, and shows congruency effects for each interference type for each experiment.

**Figure 5.1**: Congruency effects by experiment and interference type. Congruency effects were calculated by subtracting each participant’s mean reaction time for congruent trials from their mean reaction time for incongruent trials. In Experiment 3, congruency was calculated with respect to Chinese natural language metaphors (earlier events above later ones). For Experiment 4, congruency was considered both with respect to Chinese natural language metaphors and with respect to the participants newly-learned metaphors. Since Experiments 6 and 7 are identical except for the interference delay (number of trials between encoding and recalling interference stimuli), they are combined here.
5.3 Results

5.3.1 Congruency Effects over Time

To what extent are the representations examined here dynamic, changing over time? This broad question can be addressed in two ways: first, do congruency effects change over the course of a single experiment block? Second, are effects different based on when participants experienced the congruent and incongruent trials? In other words, do people who receive the congruent judgment block before the incongruent show different overall congruency effects than those who receive the reverse?

To test the first question, for each experiment I calculated the mean congruency effect for each trial (1-56), to examine relationships between trial number and congruency effect. I also repeated the process by collapsing across experiments and calculating mean congruency effects for each interference type for each trial, to see whether the trajectory of congruency effects across trials depended on the interference type someone experienced.

To examine whether block order played a role in the manifestation of congruency effects, I constructed a mixed effects model of reaction times as a function of congruency, interference type, experiment, and block order (congruent first vs. congruent second) as fully-crossed fixed effects, with participants and items as random effects.

To what extent are the representations examined here dynamic, changing over time? This broad question can be addressed in two ways: first, do congruency effects change over the course of a single experiment block? Second, are effects different based on when participants experienced the congruent and incongruent trials? In other words, do people who receive the congruent judgment block before the incongruent show different overall congruency effects than those who receive the reverse?

To test the first question, for each experiment I calculated the mean congruency
effect for each trial (1-56), to examine relationships between trial number and congruency effect. I also repeated the process by collapsing across experiments and calculating mean congruency effects for each interference type for each trial, to see whether the trajectory of congruency effects across trials depended on the interference type someone experienced.

To examine whether block order played a role in the manifestation of congruency effects, I constructed a mixed effects model of reaction times as a function of congruency, interference type, experiment, and block order (congruent first vs. congruent second) as fully-crossed fixed effects, with participants and items as random effects.

5.3.2 By Trial Number

Overall, congruency effects got smaller throughout the course of a block of judgments ($r = -0.11$, $n = 338$, $p = .05$), but this trend was present in different experiments to differing extents. Figure 5.2 shows the best fit lines for each experiments congruency effects over the course of the 56 trials.

![Figure 5.2](image.png)

**Figure 5.2:** Best fit lines for congruency effects for each experiment over the course of a judgment block (56 trials).
The overall negative correlation between trial number and congruency effect was driven by significant correlations for two experiments: Experiment 1, which included English speakers trained on new vertical spatio-temporal metaphors \((r = -0.32, p = 0.02)\), and Experiment 5, which was the first experiment to examine lateral space-time associations \((r = -0.35, p = 0.009)\). No other experiments had significant linear changes in congruency effect sizes over the course of a block \((p > .24)\). It is especially notable that while the first experiment (Exp 5) examining lateral space-time associations demonstrated a significant negative correlation between congruency effect and trial number, subsequent experiments that were nearly identical but included longer interference delays (Exps 6 & 7) did not show the same relationship. Thus, while there is an overall tendency for congruency effects to weaken over the course of a judgment block, it is unclear whether features of the particular experiment contribute to such a change over time, and if so, what those features may be.

One possible explanation for the decrease in congruency effects over the course of a block might be a practice effect: people get faster at the task over the course of a block \((r = -0.18, n = 337, p = 0.001)\), and congruency effects are likely to be smaller when reaction times overall are lower. If this were the case, we would expect a positive correlation between reaction time and congruency effect. To test this, for each experiment I found the average reaction time for each trial (1-56) and the average congruency effect for each trial. In fact, reaction times and congruency effects were negatively correlated \((r = -0.12, n = 337, p = .03)\), so an overall decrease in reaction times cannot explain the diminished congruency effects over time. Instead, it is likely that the instructions for a block of time judgments act as a separate kind of training, strengthening the instruction-consistent representation of time over the course of a block.
By Block Order

Overall, congruency effects did not differ by block order (b = 2.02, SEM = 18.63, t = -0.108, p = .91), but there was variation across experiments in the effect of block order on congruency effects, as shown in Figure 5.3. I used Experiment 1 as the baseline for comparisons since the following experiments were designed to build on and inform the initial one investigating whether learning a new system of linguistic metaphors can shape non-linguistic mental representations.

![Figure 5.3](image)

**Figure 5.3**: Congruency effects by experiment and block order.

The difference in congruency effects for the two block orders was greater in Experiment 3 (untrained Chinese-English bilinguals) than in Experiment 1 (b = 47.623.1 SEM, t = 2.1, p = .04). The participants in Experiment 3 showed especially strong congruency effects with respect to their natural language metaphors when they experienced the natural language-incongruent judgment first, followed by the congruent block. This likely reflects the role pre-existing mental representations (leading people to be faster for congruent than incongruent judgments) in addition to a practice effect,
facilitating performance on the second block as compared to the first.

A similar effect was found for the two groups of experiments in which participants made time judgments using lateral responses (for Experiment 5, one time judgment between interference encoding and recall, compared to Experiment 1, $b = 132.2 \pm 36.3$ SEM, $t = 3.64$, $p = .0003$; for Experiments 6 & 7, either 2 or 16 judgments between interference encoding and recall, $b = 83.1 \pm 18.7$, $t = 3.8$, $p = .0001$).

Overall, in cases where participants were being tested for congruency effects resulting from long-term representations, originating either in language or in cultural practices like reading and writing, those who experienced the incongruent block before the congruent block tended to show greater congruency effects than those who experienced the reverse. As mentioned previously, this is likely because people naturally become better over time, responding more quickly on the second block than the first, and this process was exacerbated by an incongruent first block, making the initial block especially difficult, and a congruent second block. It is unclear why this same block order advantage was not evident in the experiments probing congruency effects with respect to newly-learned metaphors or with respect to space-number associations, but this may be one subtle indication that those types of representations do differ in some way from those acquired through long-term experiences with space-time associations.

5.3.3 Congruency Effects & Individual Differences

Throughout this dissertation, I have compared vertical mental representations between Chinese-English bilinguals and English speakers (who may be bilingual in languages other than Chinese) since the Chinese language includes vertical metaphors for talking about time that are more systematic and productive than those in English. I have suggested that Chinese-English bilinguals in Experiments 3 and 4 demonstrated early-up biases because of their background specifically as speakers of Chinese. If this
is true, bilinguals who speak languages other than Chinese should not show the same early-up biases.

To examine whether bilinguals specifically with Chinese language experience show different early-up biases than others, I used data from the experiments that included vertical response keys (Exps 1, 3, & 4), and categorized all participants as either: a) bilinguals who spoke Chinese; b) bilinguals who spoke languages other than Chinese; c) monolingual English speakers. I constructed a linear mixed effects model to examine reaction times by language group (bilingual with Chinese experience, bilingual without Chinese experience, monolingual), experiment, judgment condition (whether the earlier key was above or below the later one), and interference type as fully-crossed fixed effects and participants and items as random effects.

Although early-up biases did not differ between Chinese-speaking bilinguals and non-Chinese-speaking bilinguals (b = 24.8 46.7 SEM, t = 0.5, p = .60), the biases also did not differ between this latter group and English monolinguals (b = -19.5 50.8 SEM, t = -0.4, p = .70). Early-up biases were marginally greater for Chinese-speaking bilinguals than for English monolinguals (b = 44.3 26.3 SEM, t = 1.7, p = .09). Note that this contrasts with the significant difference reported for Chinese-English bilinguals and English speakers in Ch. 3, likely because splitting the English speaker group into subgroups (non-Chinese bilinguals and English monolinguals) decreased the sample size, and therefore the power of the analysis.

Further, examining each of the three language background groups alone revealed that while Chinese-speaking bilinguals did show systematic early-up biases (b = 78.1 32.2 SEM, t = 2.4, p = .016), neither non-Chinese-speaking bilinguals (b = 50.9 53.0 SEM, t = 1.0, p= .34) nor English monolinguals did (b = 18.6 22.5 SEM, t = 0.8, p = .41). Earlier-is-up biases for each language background group are shown in Figure 5.4.
Figure 5.4: Earlier-is-up biases for Chinese-English bilinguals, bilinguals in languages other than Chinese, and English monolinguals. Using Chinese natural language metaphors as a starting point, Congruent = earlier is up and Incongruent = earlier is down.

5.3.4 Congruency Effects by Stimuli Features

An additional way we can learn more about the patterns in space-time and space-number associations is by looking at how features of the stimuli alter congruency effects. By doing so, we can ask whether space-time associations demonstrate metric properties. For instance, are events that are close in time also represented in close proximity in our minds, compared to events that are more spaced out over time? The images that participants made temporal judgments about varied in the time spans they depicted, so we can examine congruency effects as a function of how great the time lapse depicted by the stimuli was.
We can also probe the nature of these space-time and space-number associations by examining whether they manifest differently for people with different verbal and spatial working memory capabilities. Assuming that the pre-experiment calibration truly resulted in interference conditions that taxed individuals’ working memory resources comparably, we can address this question by comparing congruency effects for people who completed the task with a low interference load (reflecting lower working memory capacities) and those who complete it under a higher load (who presumably have stronger working memory capacities). This analysis can open doors for understanding the relationship between working memory abilities and systematic space-time and space-number associations.

**By Temporal Distance from Present**

For these analyses, we rank ordered the stimuli by the temporal duration they expressed. The 56 sets of images were rank ordered such that #1 reflected a very short time lapse from one image to the next within the sequence and #56 reflected a much greater temporal distance among the images. For example, a sequence of images that showed a glass of milk being filled reflected a short time lapse between events, while images of a person aging from a child to an older adult reflected a much longer time lapse.

A mixed effects model of reaction times that included stimulus temporal rank, congruency, interference type, and experiment as fully crossed fixed effects revealed no significant effect of temporal rank on congruency effects, as there was no interaction between congruency and temporal rank ($b = 0.83 \pm 1.2$ SEM, $t = 0.7$, $p = .49$). Overall reaction time was also not affected by stimulus temporal rank ($b = 0.45 \pm 1.6$ SEM, $t = 0.3$, $p = .78$). Figure 5.5 shows best fit lines for congruency effects for stimuli ranging from short to long temporal durations in each experiment. There were no differences in
the effects of stimulus duration on congruency effects for the different interference types (ps > .75).

**Figure 5.5**: Congruency effects by stimulus duration for each experiment. Stimulus duration was ranked so that sequences with the shortest time lapse between images are at the low end of the scale and sequences with longer time lapses between images are at the high end.

**By Number of Interference Distractors**

To examine the effect of a person’s working memory capacity (operationalized by the number of interference distractors the participant received after calibration) on strength of space-time associations, I included data from all experiments in which participants completed one judgment between interference coding and recall (Exps 1, 3, 4, & 5). I included data for the interference levels that at least 25 participants had experienced, which included 3-6 distractors. This sample included 288 participants and represented a total of 83% of all participants who completed a time judgment task in one of the aforementioned experiments. Interference distractors were forward-difference coded so that comparisons were made between subsequent levels.
The mixed effects model included the fully-crossed fixed effects of number of distractors, experiment, interference type (including only spatial and verbal interference), and congruency. An interaction between distractors (3 vs. 4) and congruency revealed that participants who had a load of 4 distractors, regardless of whether they were spatial or verbal, showed greater congruency effects (with respect to either newly-learned linguistic metaphors, long-term natural language metaphors, or cultural practices, depending on the experiment) than participants with 3 distractors (b = -154.0, SEM = 66.1, t = -2.3, p = .02). There was similarly an interaction between distractors (4 vs. 5) and congruency, but in the reverse direction, demonstrating that participants with 5 distractors showed smaller congruency effects than those with 4 (b = -192.7, SEM = 59.2, t = 3.3, p = .001). There was no difference in congruency effects for participants who saw 5 vs. 6 distractors (b = 35.8, SEM = 91.1, t = 0.39, p = .69). Figure 5.6 shows congruency effects across the different levels of interference distractors for verbal and spatial interference. Despite
significant interactions between congruency effects and number distractors (3 vs 4 and 4 vs 5) the overall trend suggests that people who underwent different interference loads did not show representations that differed systematically.

### 5.4 Discussion

One advantage of using overlapping methods for six of the experiments in this dissertation is that the work yields a large dataset that can be examined for overarching patterns that may not emerge in individual experiments, or may emerge in subsets of experiments. The goal of the meta-analyses presented in this chapter is to shed light on patterns in the mental representations described in the preceding chapters, such as how the representations manifest on different time scales, the role of an individual’s language background on the representations, and what features of the stimuli influence their manifestation.

The first topic explored was whether systematic space-time or space-number effects manifest differently over time. I first examined this by considering changes in congruency effects over the course of a 56-judgment block. On incongruent blocks, participants were asked to use key mappings that were incongruent with relevant associations (whether newly learned or long established). This instruction can itself be viewed as a kind of training – creating a new association between space and time. If the response instruction is indeed providing a kind of counter-training, we might expect that the congruency effect should be stronger in earlier trials and weaker in later trials, decreasing as the number of counter-training trials increases. This is indeed what was found overall, which sheds light on the dynamic nature of our systematic associations between space and abstract concepts like time or number.

I also examined another kind of timing variable on the manifestation of systematic
representations; namely, the order in which participants experienced the congruent and incongruent blocks. Overall, block order did not affect the size of congruency effects, but a comparison of the effect of block order on congruency effects across experiments revealed that for a subset of experiments, block order did have an effect on congruency effects. Specifically, in cases where participants were tested for congruency effects resulting from long-term representations, originating either in language or in cultural practices like reading and writing (Exps 3, 5, 6, & 7), those who experienced the incongruent block before the congruent block tended to show greater congruency effects than those who experienced the reverse, which likely reflects a combination of a congruency effect and a practice effect. However, in Experiments 1 and 4, where participants demonstrated congruency effects with respect to newly-learned linguistic metaphors, the same relationship between block order and congruency effect did not hold. Future research might work to understand whether there is something different in the nature of representations arising from a newly-learned convention rather than a long-term one, as this finding seems to hint at.

This meta-analysis also allowed me to rule out the possibility that Chinese-English bilinguals showed vertical biases in Experiments 3 and 4 as a result of their general bilingualism, as opposed to specific Chinese-English bilingual experiences. A comparison of early-up biases across the experiments with vertical responses revealed that participants who were bilingual in a language other than Chinese did not differ in their early-up biases from English monolinguals. This finding suggests that Chinese language experience in particular is a crucial feature of the participants who showed systematic early-up biases, and lends support to the possibility that experience with Chinese natural language metaphors that place earlier events above later ones can cause Chinese speakers to think about time consistently.

The examination of the role of an individual’s working memory capacity did not
suggest that this particular individual difference plays a crucial role in the manifestation of specific space-time or space-number associations. The intention behind examining the effect of interference load on congruency effects was that these associations might be different in nature for people who have low vs. high working memory (particularly spatial working memory) capacities. This doesn't seem to be the case, but future work might include a within-subjects manipulation of number of interference distractors to more directly address the role of working memory load on the manifestation of congruency effects.

Similarly, an analysis of the role that the time lapse depicted by the stimuli might play in the manifestation of systematic space-time associations was inconclusive. Prior work has suggested that mental representations of number do have metric properties, as demonstrated by spatial-numeric association of response codes (SNARC) effects: the farther a number is from the midpoint relevant to a particular task, the greater the congruency effect (e.g. [34]). Although the particular task and stimuli used in the experiments described here did not show a similar metric property for space-time associations, it is possible that different methods may. Thus, future work should work to confirm or rule out the possibility that our mental representations of time exhibit distance effects as mental representations of number do.

Up to this point, this dissertation has systematically examined mental representations of time (and number), revealing first that participants show congruency effects on a non-linguistic space-time association task, reflecting representations consistent with newly-learned metaphors, natural language metaphors, and cultural practices. Further, the experiments presented here have repeatedly shown that these systematic space-time (or space-number) associations are unaffected by the addition of interference. For the experiments that addressed congruency effects with respect to newly-learned or long-term natural language metaphors, verbal interference in particular did not alter congruency
effects, suggesting that language can create representations that are non-linguistic in nature.

The following chapter builds on this work by investigating a different kind of non-linguistic evidence of a representation; namely, through gestures. The final chapter turns the tables to investigate the emergence of a system of linguistic metaphors to reflect a previously-existing mental representation.
Chapter 6

New Space-Time Associations in the Body

6.1 Abstract

Chapter 2 (Exp. 1) demonstrated that learning a new way to talk about time (using vertical metaphors) creates new non-linguistic representations of time that are consistent with newly-learned metaphors. Those prior conclusions were based on performance on a non-linguistic space-time association task, which, while methodologically precise, forces responses that are spatialized on the axis of interest. In this chapter, we ask whether mental representations of time acquired through new linguistic metaphors can manifest spontaneously. Specifically, do we see signs of newly-learned metaphors in people’s co-speech gestures? To address this question, we taught participants the same systems of metaphors as in Exp. 1 (that earlier events are above or below later ones). After learning a new system of metaphors, participants then described temporal concepts for an ostensibly non-native English speaker, and we later analyzed their gestures. We focused specifically on whether gestures were consistent with newly-learned metaphors,
to understand whether mental timelines created by learning patterns in language can manifest spontaneously.

6.2 Introduction

The guiding question of this dissertation has been on how linguistic metaphors shape the way we mentally represent complex concepts. More specifically, it has focused on whether learning new metaphors to talk about time leads to new mental representations, and what the nature of those representations might be. Chapter 2 (Exp. 1) shows that after learning that earlier events are either above or below later ones, participants showed new metaphor-consistent representations on a non-linguistic space-time association task. The subsequent experiments in this dissertation built on the paradigm used in the first experiment to further probe the nature of the effects of metaphor on thinking about time, by including verbal and spatial interference, comparing timelines on multiple axes, in groups with different long- and short-term experiences, and considering an additional spatial representation – number.

While the non-linguistic space-time association task used in the prior experiments offers methodological control and precision, it also forces spatialization on the axis of interest. For example, metaphor-consistent vertical representations emerged in Exp. 1 in a task in which only vertical representations (or no consistent representations) could possibly emerge. As a result, in this chapter we ask whether newly-learned metaphors can shape mental representations in ways that are evident in spontaneous behaviors, particularly when the axis of interest is not built into the task.
6.2.1 Gestures

Many cognitive processes are difficult to observe directly, but gesture – especially co-speech gesture – offers one window into thought [85, 55]. Gestures often reflect our thoughts, including ones that aren’t verbalized [55]. Because gesture frequently accompanies speech, it provides insight into the mental representations people have in mind while communicating [45].

Another advantage of gesture is that it does not force a particular spatialization [124], as opposed to the methods used in the prior experiments presented in this dissertation. Do newly learned metaphors shape mental representations of time even when the task at hand is not constrained to a particular spatialization? Gesture is uniquely suited to address this question because it allows representation on three axes, independently and simultaneously, so that in theory all of the space in front of the body (and some space behind) is available [124].

**Gestures reveal mental timelines on the lateral axis**

Prior work has revealed that when talking about time, English speakers tend to gesture laterally, placing earlier events to the left of later ones [27, 29], particularly when they are not prompted to be deliberate in their gestures [20]. Although language use does not reflect these mental timelines (but see Ch. 7), their emergence in gesture mirrors their emergence in cognitive psychology experiments [120, 126, 43, 87, 125] (also, Ch. 4, this dissertation).

**Gestures reveal mental timelines consistent with language**

Although lateral gestures reflect mental timelines not present in language, other work has shown that gestures are often consistent with the language they accompany. For example, Aymara speakers refer to the past as in front and future as behind them...
in language and in gesture [94], in contrast to bilinguals from the same area who have experience with Spanish. Aymara gestures also provided insight that language alone did not. For example, the researchers sometimes observed Aymara-consistent gestures accompanying Spanish speech, suggesting that experience with Aymara created a front-to-back mental timeline for people that existed even when speaking a language that didn’t make use of such metaphors. These participants also made gestures along the left-to-right axis, again revealing mental representations that were absent from spoken language.

Other work has demonstrated that English speakers sometimes gesture consistently with English metaphors for time (with the past in back and future in front), particularly when they’re asked to deliberately gesture [20]. Since sagittal gestures were less frequent when participants were not asked to intentionally gesture, this work suggests that the deictic language, which often makes use of sagittal spatio-temporal metaphors, may encourage people to think about time as flowing along the sagittal axis in a way consistent with the linguistic metaphors, as reflected in their patterns of gesture.

Chinese speakers’ temporal gestures have also been noted to reflect Chinese linguistic metaphors [59, 57, 58]. Participants in previous experiments gestured vertically when talking about time, consistent with Chinese vertical metaphors (i.e., literally, above week or below week). Notably, they gestured vertically more often when the time concept they were discussing contained a vertical linguistic metaphor than when it did not. In addition, these same participants gestured vertically more when describing those terms that included vertical linguistic metaphors in Chinese, than when they described the English translations of those terms in English. These two results suggest that linguistic metaphors can activate vertical mental metaphors, as reflected in gestures, as participants are talking. Although vertical mental timelines may be present in Chinese-English bilinguals when talking about non-metaphorical (or sagittal-metaphorical) terms and when talking in English, vertical mental timelines are more likely to manifest in the
presence of consistent language.

**Gestures reveal dynamic mental timelines**

The work discussed so far demonstrates that gestures can reflect mental representations of time, whether they are explicit in language or not. The work also begins to show that mental timelines are dynamic, with different representations often used in different circumstances, for example with lateral gestures dominating implicit co-speech gestures and sagittal gestures dominating directly elicited gestures [20].

The dynamic nature of mental timelines is also evident in the work on Chinese-English bilinguals’ co-speech gestures for talking about time [59, 57, 58]. For example, linguistic choices (whether the term contained a vertical metaphor) influenced people’s tendency to gesture vertically. Additionally, most participants gestured along multiple axes while explaining the temporal concepts, and some gestured along all three at different times. During a single experimental session, participants seem to rely on multiple timelines at different times.

Additional work shows that not only do people switch between axes when thinking about time, but they also systematically combine multiple timelines [124]. When describing temporal concepts, participants in this study gestured along the sagittal and lateral axis simultaneously, producing doubly congruent gestures (those that refer to the future as to the front and right or past as to the back and left) more often than singly congruent ones (gestures that use both axes but refer to the future as either to the front or right only or the past as to the back or left only). These gestures shed light on the combination of mental timelines specifically, instead of some mechanical advantage for front-right and back-left gestures, since this pattern only held for describing temporal, and not spatial concepts.

As with cross-linguistic work that reveals consistencies in how people talk and
think about time, the existing gesture work cannot verify that linguistic metaphor in particular, without influences from artifacts, cultural practices, and social experiences, can create the kind of representations that are revealed in spontaneous co-speech gesture. This experiment seeks to do just that.

Taking inspiration from prior work that has examined gesture patterns in natural language speakers, in the current work we examined gesture patterns in speakers who had learned a new system of vertical metaphors for talking about time in the lab. After learning the new metaphors, participants described time concepts. We video-recorded their responses, and analyzed their gestures, along with their speech, to examine whether representations from the newly-learned metaphors manifested in their spontaneous gestures. In short, we asked: After learning new vertical metaphors for talking about time, will people’s gestures reflect those newly learned associations?

6.3 Methods

6.3.1 Participants

Thirty-eight undergraduates at UC San Diego participated for course credit. One participant was excluded for producing no gestures, and two others were excluded for correctly reporting the purpose of the experiment during debriefing.

6.3.2 Materials & Procedure

Metaphor Training

In the first part of the experiment, participants learned a new system of metaphors for talking about time. This portion was identical to the methods used in Exp. 1 (Ch. 2).
Spontaneous Gesture Elicitation

After learning the new system of metaphors, the experimenter told the participant there was a second experiment to complete. They went to another room on a different floor of the building. Following prior work [124], to get participants warmed up to speaking (and gesturing), participants first described how they got to the building where the experiment was held.

The main task was modeled after prior work that examined co-speech gestures when people talked about time [124]. Participants were instructed to describe different words to someone who knew very little English. Since that person wasn’t present, their responses would be videotaped. Participants sat on a stool across from the experimenter and the computer that taped their responses and displayed the words they were to describe. They were told that there were 18 words total, each would stay on the screen for 25 seconds, and they should try to continue explaining for the entire 25 seconds.

We used a subset of the stimuli used by Walker & Cooperrider [124]. They included: (a) temporal words: *tomorrow, yesterday, earlier, later, last year, next month, past, future*; (b) spatial words: *near, far, front, behind, above, below*; (c) abstract words: *idea, success, fear, happiness*.

These stimuli were pseudorandomized to ensure that the temporal words were relatively spread out throughout the experiment. Each participant saw one of two stimulus order versions – the original pseudorandomized version or its mirror image. The words auto-advanced after 25 seconds.

At the end, participants completed a survey asking them what they thought the purpose of the experiment was, if they believed they gestured (and if so when), and demographic information, including languages they spoke. Before they left, the experimenter explained the connection between the metaphor training and the word description tasks.
6.3.3 Gesture Analyses

Three coders, all blind to participant condition, contributed to the gesture analyses. We used ELAN annotation software (Max Planck Institute for Psycholinguistics, Available at: http://www.mpi.nl/tools/elan.html). The first coder identified all descriptions of temporal words in the videos. The second and third coders watched the temporal segments and documented each movement (by either hand, head, or whole body). Since agreement between the two coders was high for a random subset of the videos (73.5%), the first coder’s annotations were used for analyses.

Each gesture was coded for movement along 3 axes (to the best extent possible using a 2-D video). We considered each gesture as a vector whose direction was based on the ending position relative to starting position. We coded the timeline each gesture implied. For example, if the word tomorrow was accompanied by a rightward movement, this gesture was coded as earlier concept to the left of later. The word yesterday, accompanied by a leftward movement would similarly be coded as earlier concept to the left of later. We used this convention for all 3 axes. Many gestures were coded for displacement along multiple axes. Examples are shown in Figures 6.1 and 6.2.

We used a logistic regression to analyze whether participants’ gestures contained a vertical displacement that was congruent or incongruent with their newly-learned metaphors. For every gesture with vertical displacement, timeline was coded as 1 if the gesture reflected a bottom-up timeline and 0 if it reflected a top-down timeline. Participants were included as a random effect (items could not be included as a random effect because not all participants produced the minimum of one vertical gesture for each word prompt).
Figure 6.1: The participant’s arms move down and to the right when she says not now, the next day, referring to tomorrow. Since compared to now, tomorrow is later, her gestures were coded as reflecting a) the earlier event is above later, and b) the earlier event is to the left of later. The blue line indicates the direction of the flow of time in the timeline applied by the participants gestures. It does not reflect the direction of movement.

6.4 Results

Overall, participants were numerically, but not statistically, more likely to produce gestures with a vertical displacement congruent with their newly learned metaphors (53.9%) than incongruent (b = -0.37, SE = 0.28, p = .19). Figure 5.3 shows the proportion of gestures with a vertical displacement that were both congruent and incongruent with the newly learned metaphors for both training groups.

6.4.1 Item Analyses

We divided our items into 4 sets of complementary pairs (earlier/later, past/future, last month/next year, tomorrow/yesterday) to examine whether tendencies to make
Figure 6.2: This gesture was coded as: a) the earlier event is above later b) the earlier event is to the left of later, since the participants arms move up and to the left when describing before, as compared to now; and c) earlier event is behind later, since her arms move backwards as she describes before. Thus, this gesture was coded as showing movement on all three axes. The blue line indicates the direction of the flow of time in the timeline applied by the participants gestures. It does not reflect the direction of movement.

gestures whose vertical displacement was congruent with newly learned metaphors differed across the terms people were describing. Proportions of gestures depicting earlier events as higher than later ones (the more common tendency) for each pair of stimuli and each training group are shown in Figure 5.4.

Notably, the only pair that did not show new metaphor-consistent gesture patterns was past/future. On the other hand, the pair that led to the numerically greatest new metaphor-consistent gesture patterns was earlier/later. This is notable because the metaphor training only referred to sequential time (two events in relation to each other [earlier/later], like breakfast/dinner), and did not refer to past/future deictic time (one event in relation to the present). It is possible that participants did not generalize the relationships between space and time as specified by their newly learned metaphors to
deictic time, since these metaphors were never used to express deictic relations. Repeating the main analysis (logistic regression) to examine a congruency effect between newly-learned metaphors and gestures vertical displacement without including the past and future word prompts revealed that participants were marginally more likely to gesture in metaphor-congruent ways (56.3%) than incongruent (b = -0.53, SE = 0.31, p = .09). This inconclusive, but suggestive finding renders this a promising pilot study.

6.4.2 Gestures on the Sagittal Axis

Most gestures (83%) reflected a back-to-front timeline on the sagittal axis. This finding is consistent with deictic metaphors in English, which refer to past events as behind the body and future events as ahead. The metaphors that participants learned in the lab did not affect their gestures on this axis (b = 0.22, SE = 0.61, p = .71).
6.4.3 Gestures on the Lateral Axis

We expected our participants, all of whom are English speakers, to gesture about time as flowing from left to right, as observed in prior work [20, 124]. Indeed, 86.4% of all gestures with lateral displacement reflected this timeline. We then compared patterns in gestures on the lateral axis for participants in the two metaphor training groups. Since the metaphor training included only vertical linguistic metaphors, there was no reason to expect lateral gesture differences between the two groups, and they did not differ in their use of the left-right axis for gestures (b = -0.73, SE = 0.63, p = .25).

One participant in the group that learned that earlier events are below later ones was a native Farsi speaker (self-rated proficiency as 5/5) who also spoke Arabic. Prior work suggests that people who read and write languages like Farsi and Arabic (and
Hebrew) mentally represent time from right to left in contrast with English speakers’ left-to-right mental timelines [120, 81, 43]. This one participant did indeed produce more gestures that reflected a right-to-left mental timeline (21/28 vertical gestures, 75%) than left-to-right.

6.5 Discussion

In this experiment, we asked: does learning a new system of metaphors for talking about time shape thought in a way that will manifest in spontaneous co-speech gestures? American university students learned new vertical metaphors for time, which placed earlier events either above or below later ones. In the second part, which was disguised as a separate experiment, participants described the meaning of temporal concepts.

The data from this pilot study suggest that after learning new systems of linguistic metaphors, participants produce gestures with vertical displacements consistent with their newly learned metaphors. Their gestures reflected the canonical left-to-right timeline and hinted at the timelines implied by the new metaphor systems they encountered in the lab, particularly when the temporal concepts being described were the same ones the newly learned metaphors applied to. This finding suggests that people may integrate new representations from language into their existing representations for thinking about time, which is consistent with earlier work showing that mental timelines on the lateral and sagittal axes can be co-activated and reflected in peoples co-speech gestures [124].

It is noteworthy that the participants in this experiment produced gestures with very subtle vertical displacements, especially compared to their movements along the lateral and sagittal axes. In fact, we did not identify any gestures that had a vertical component and didnt have movement along one of the other axes. This finding helps assure (in conjunction with participants responses at the end of the experiment about
what they thought we were investigating) that they were not intentionally gesturing in line with the new metaphor systems they had learned. Instead, the new vertical metaphors seemed to have a subtle influence on thought, meshing with other spatial representations that have presumably been reinforced throughout participants’ lifetimes.

Because the main effect of interest – consistency of vertical metaphors with newly learned metaphors – was only marginally consistent, this experiment should be considered as pilot work and should be replicated. Future work using quantitative measurements of gestures, rather than 2-D videotaping with human coders, can increase the precision of measures of subtle gestures. This method would also allow further investigation into potential metric properties of temporal gestures: For example, does the hand move farther for last year than it does for yesterday? Nuanced analyses like these are difficult without quantitative movement measurements.

6.5.1 Mental Timelines are Dynamic

The data presented here also contribute to a growing body of work demonstrating that multiple timelines can coexist in our minds [13, 44, 87, 65, 58]. Our representations of abstract domains like time are not fully articulated, logically cohesive knowledge structures, but rather a bricolage of many different (sometimes conflicting) structures that are brought to mind for different purposes. Inspired by the metaphor of mental spatial representations as a cognitive collage [119], they have been described as ad hoc: always incomplete and schematic, containing no more detail or metric precision than is required by the context ([20] p. 666).

Our data support an emerging view that metaphorical cognition may more be more fruitfully understood as a phenomenon that emerges from a dynamic system [51, 124]. The dynamical view of metaphor takes constraints from historical and cultural knowledge, to people and environments as coupled systems, to the fast firing of neurons
into account to understand a single figurative experience [51]. Considering our data in this more encompassing framework of metaphor, it makes sense that we observed gestures along multiple axes simultaneously, and that prior experiences, in the lab (learning new metaphors) and outside (extensive experience with artifacts, cultural practices, and linguistic metaphors), acted as constraints as participants thought and spoke about the temporal concepts in our study.

6.5.2 Linguistic Metaphors Shapes Gestures

In this experiment, we set out to understand whether linguistic metaphor can shape mental representations of time in ways that can be seen in spontaneous behavior. We showed that after learning new vertical metaphors to talk about time, participants’ spontaneous co-speech gestures revealed subtle influences of their newly learned metaphors. Our findings suggest that the vertical mental timelines implied by the linguistic metaphors that participants learned were integrated into the dynamic representations they relied on for thinking about time. The experience learning the new metaphors was one of many that contributed to internal mental representations and the spontaneous demonstrations of those representations through gesture.

The role of many experiences, such as learning new metaphors in the lab, extensive experience reading and writing, observing others’ gestures, using natural language spatio-temporal metaphors, physically moving through the world, and the current environment, can likely all shape the way people think about time in any given moment. Although mental representations of time have been studied in more detail than many other abstract mental representations, it is likely that this broad view that a multitude of constraints influence metaphorical thinking in real time generalizes to our cognition about many other ideas. Future work should continue to uncover the factors that contribute to metaphorical thought and their respective roles in how we think about the incredibly
complex world around us.
Chapter 7

Do metaphors move from mind to mouth? Evidence from a new system of linguistic metaphors for time

7.1 Abstract

Languages around the world use a recurring strategy to discuss abstract concepts: describe them metaphorically, borrowing language from more concrete domains. We plan ahead to the future, count up to higher numbers, and warm to new friends. One influential account says that these systematic linguistic metaphors originated in shared, non-linguistic ways of thinking – for instance, conceptualizing time as if it were a back-to-front timeline. On the other hand, systems of linguistic metaphors could be the cause – not the effect – of shared conceptualizations. Here, we present a case study of a variety of American English in which a shared, non-linguistic conceptualization of time has become crystallized as a new system of linguistic metaphors. Speakers of various languages, including English, conceptualize time as a lateral timeline, with the future rightward and
the past leftward. Until now, this conceptualization has been thought to be absent from speech in any language. In two studies, we document how members of the US military, but not US civilians, talk about time using conventionalized lateral metaphors (e.g., move the meeting right to mean move the meeting later). We argue that, under the right cultural circumstances, implicit mental representations become conventionalized metaphors in language. ¹

7.2 Introduction

In English, we plan ahead to the future, count down to lower numbers, and warm to new friends. These phrases exemplify a strategy that recurs in languages around the world: abstract concepts are described metaphorically using words and phrases whose meanings, originally and primarily, are more concrete [74, 72]. These metaphorical expressions are frequently more than one-off turns-of-phrase. We dont just plan ahead, but can also look back on the past, approach a deadline, and experience time flying by. In these and countless other expressions, the domain of Space is used systematically to discuss the domain of Time, with specific, recurring associations – for instance, between length and duration (“a long day”) or front/back location and past/future (“look ahead to the future”). These metaphorical expressions for Time, therefore, form a cohesive system – and systems of metaphorical expressions exist for other domains, too. Here, we investigate the historical origin of systematic, conventional metaphors in language. Where do these metaphor systems come from?

One influential approach says that these systematic expressions are the linguistic manifestation of pre-existing systems in thought. On this thought-first account, linguistic communities begin by conceptualizing some abstract domain by mapping it to another, ¹Hendricks, Bergen, & Marghetis. under review at Cognitive Science.
often more concrete, domain. Perhaps they conceptualize time as sagittal space; or affection as physical warmth; or power as vertical location. These shared, implicit conceptualizations might themselves have various origins, ranging from innate biases (e.g., [17, 19, 78, 110]), to the morphology and kinematics of our bodies (e.g., [20]), to recurring experiences (e.g., [75, 18]). Regardless of their origin, the critical premise of this thought-first approach is that metaphorical mappings between two domains begin in the mind and only later become conventionalized in language. Said otherwise, linguistic metaphors are historical byproducts of shared, implicit, nonlinguistic conceptual metaphors [74, 112, 49, 37]. Versions of this general account have been put forward in various disciplines, including cognitive psychology, historical semantics [118, 74], and developmental psychology [28].

Some of the best evidence for the thought-first account is that systems of linguistic metaphor typically have analogs in non-linguistic thought. For example, in English, importance is described in terms of heaviness (e.g., *weighty topics, heavy conversations, dense papers*). But people also think about importance in terms of heaviness. In one experiment, participants who held heavier (vs. lighter) objects subsequently gave higher ratings of job candidates [1], monetary value [70], and the seriousness of diseases [71] as if the experience of physical heaviness primed, implicitly, the more abstract concept of importance. Similarly, English speakers talk about the past and future as if they are back and ahead, and they also think in that way: they lean forward when thinking about the future [86], gesture forward spontaneously when talking about the future [20], and are faster to move forward in response to future-related stimuli [117, 104, 121, 102]. Thus, systems of linguistic metaphors are typically accompanied by analogous systems of thought, exactly as we would predict if linguistic metaphors are the product of preexisting conceptual metaphors.

However, there are good reasons to suspect that systematic linguistic metaphors
might be the cause, rather than the effect, of non-linguistic conceptual metaphors. Indeed, habitually using new, systematic metaphorical expressions is known to create new habits of metaphorical thought. For example, Dutch participants were taught to talk about musical pitch using new spatial metaphors (e.g., thick pitch for low pitch, thin pitch for high pitch). Afterwards, the task-irrelevant thickness of visual stimuli had a systematic effect on their pitch perception, suggesting that learning to talk about pitch in terms of physical thickness caused them to think about pitch as if it were physical thickness [36]. Similarly, after English speakers were taught to talk about temporal sequences as if they were arrayed vertically (e.g., You eat lunch above [after] you eat breakfast), they began to think about temporal sequences along the vertical axis (Experiment 1 [65]. Learning a new system of linguistic metaphors, therefore, can create new conceptual metaphors in speakers’ minds.

This causal impact of language on thought suggests a different interpretation of the widespread co-occurrence of metaphorical language and metaphorical thought. Metaphorical systems may actually originate in language – perhaps as idiosyncratic rhetorical flourishes, perhaps as creative responses to communicative demands, perhaps due to fortuitous speech errors that conflate domains that are coupled systematically (e.g., differences in power are often accompanied by differences in height). Once such expressions become conventionalized within a language community, people may use the relations implied by the language to structure their mental representations. Shared conceptual metaphors, therefore, might be the effect, not the cause, of conventional linguistic metaphors – in short, a language-first account of the co-occurrence of community-wide metaphorical language and thought. This possibility is consistent with extensive work on linguistic relativity, supporting the notion that features of natural languages can systematically shape the way speakers think (for review, see [129]).

Indeed, there is a surprising dearth of direct evidence for the thought-first account
of the origins of systematic metaphorical expressions. The historical record, for instance, is replete with semantic changes that appear, in retrospect, to have been driven by a shared, implicit conceptualization [112, 118]. Lexical items that are originally reserved to describe perceptual experience (\textit{rough surface}) are gradually extended to more abstract senses (\textit{rough experience}) in ways that suggest an underlying metaphorical conceptualization (e.g., [112, 130]). But this diachronic evidence cannot adjudicate between \textit{thought-first} and \textit{language-first} accounts, since it is impossible to measure directly the implicit conceptual representations of historical language communities. A more direct demonstration of the \textit{thought-first} account would involve synchronic evidence of the emergence of a novel system of metaphorical expressions, within a language community that is known to already conceptualize the domain in the same way. Unfortunately, the pervasiveness of metaphors that are both linguistic and conceptual makes it difficult to capture, in real-time, the emergence of a new system of metaphors. For instance, while there is substantial evidence that communities think metaphorically about time (as space) and affection (as warmth), these communities already speak using corresponding linguistic metaphors, too.

Thus, we set out to find an instance that filled in this evidential gap – a case study in the historical relation between metaphorical thought and language. We sought an instance where members of a linguistic community were known to think about a domain metaphorically, but where that metaphorical construal was only now becoming crystallized in speech. This would serve as a proof-of-concept of the \textit{thought-first} account that a metaphorical construal in \textit{thought} can emerge eventually as a new system of metaphorical expressions in \textit{language}. By buttressing the \textit{thought-first} account, such a case study would have implications for theorizing about the historical emergence and contemporary distribution of linguistic metaphors. For instance, it would help explain the ubiquitous co-occurrence of metaphorical systems in language and thought. And it
would support an account of the cross-linguistic pervasiveness of certain metaphors: if recurring experiences or innate biases produce cross-culturally shared ways of thinking metaphorically, then these should also result in cross-culturally shared ways of talking metaphorically [28, 74, 49].

Here, we describe just such a case study: an instance of an established conceptual metaphor that preceded the emergence of a conventionalized system of linguistic metaphors. Speakers of American English are known to conceptualize time as a left-to-right path (a mental timeline; e.g., [120]) but never speak about time using the language of lateral space (e.g., can we move the deadline to the left [/right]?). Here, we document a community that has conventionalized the use of a system of lateral metaphors for time: members of the U.S. military.

Below, we begin by reviewing what is currently known about spatial metaphors for time in both language and thought. We then give some context for our case study of the U.S. military, before turning to our empirical studies of their system of lateral (left/right) metaphors for time.

### 7.2.1 The Coupling of Language and Thought about Time

Within a given culture or community, individuals think and talk about time in ways that are both stable and shared. This often involves using space to structure both temporal speech and temporal understanding (for reviews, see [12, 93, 128]).

English speakers, in particular, speak about time as though it were represented along the sagittal (front-back) axis. The future is ahead. The past is behind. You can look forward to an event in the future, and think back to the past. This system of metaphorical expressions is both conventionalized and highly productive, so that novel, unusual sagittal metaphors are readily understood by native speakers (e.g., Don’t just look ahead, passively, toward your future – get on a motorcycle and race toward your goals!).
The way English speakers talk about time, moreover, aligns with how they think about time. When they make decisions about events, they are faster to respond to future events by moving forward, and faster to respond to past events by moving backward [104, 121, 102]. They are faster to make time judgments when future-related words are shown in front of an image of a person and past words behind [117]. When imagining the future, people lean forward, and when thinking about the past, they lean back [86]. And they gesture forward when talking about the future, but backwards when talking about the past [20]. Multiple sources of evidence, therefore, suggest that English speakers not only speak but also conceptualize time along the back-to-front sagittal axis.

Similar systems of spatial metaphors have been documented in many languages around the world [93]. Most follow the same pattern as English, with the past associated with space behind the body, and the future with the space in front. Some languages, however, deploy other conventions. Aymara, an indigenous language spoken in the Chilean Andes, also uses the sagittal axis, but reverses the English mapping: past events are in front and future events behind [94]. Mandarin Chinese uses vertical terms (up/down) to describe temporal sequences: earlier events are up and later events are down [13, 44, 87, 131]. For speakers of both Aymara and Mandarin, moreover, the spatial construal deployed by their language is also evident in their non-linguistic thought. Speakers of Aymara, for instance, gesture in ways that are congruent with their sagittal spatial metaphors, even when they are not using the metaphors in speech [94]. Speakers of Mandarin think about temporal sequences using the vertical axis [87, 59, 131], even during non-linguistic tasks [10, 13, 44]. Thus, in linguistic communities around the world, there is often a tight coupling between temporal language and temporal thought [93, 128].

But there are exceptions. In addition to a sagittal (front-back) construal of time, speakers of English and other languages also conceptualize time as a left-to-right path.
For instance, when asked to arrange physical depictions of sequences of events, they arrange them from left to right [120, 14, 116]. During natural speech, they gesture to the left for earlier events and to the right for later ones [29, 20]. And English speakers are faster to indicate that one event occurred earlier than another by responding on their left side, but faster for later events when responding on their right [43, 87, 126, 125]. In the minds of English speakers, therefore, the past and future are not just behind and ahead, but also to the left and right.

This shared conceptualization of time, however, is absent from the English language itself. English speakers can say that they look back on the past, but they would never say, in standard speech, that they look to the left. Indeed, while a left-right conceptualization of time is thought to be widespread in literate cultures [93], there is no documented language that talks about time using the lateral axis for example, with the past to the left and the future to right. This is thus a linguistic gap, where a widespread, stable, shared conceptual metaphor has not yet been conventionalized as a system of linguistic metaphors. According to thought-first accounts of the evolution of metaphorical language, therefore, this is a metaphor that is ripe to make the leap from mind to mouth.

### 7.2.2 Case Study: The U.S. Military

Anecdotal evidence from members of the authors’ families suggests that one community of English speakers has, in fact, started to use left-right metaphors when talking about time: members of the United States military. According to these anecdotes, members of the US military can reschedule a meeting to an earlier [later] time by asking to move it to the left [/right]. These and similar expressions were not seen as poetic or unusual, but as a standard way to talk about time.

This appears to be an established, but still emerging, phenomenon. For instance,
the US military has two distinct types of members: Officers, who are generally college-educated and in charge of leadership and planning; and Enlisted members, who are responsible for implementing Officers’ plans. According to anecdote, these lateral spatial metaphors are especially pronounced among Officers, compared to Enlisted members. Conversations with members of the US military, moreover, suggested that this novel system of linguistic metaphors is already governed by dialect-specific conventions for instance, a preference for dynamic descriptions (Fridays meeting was moved to the left, to Wednesday) rather than static descriptions of temporal relations (Wednesdays meeting is to the left of Fridays meeting).

This linguistic subgroup’s use of lateral metaphors could thus provide an opportunity to document a process that has been posited by thought-first accounts: the emergence of a new system of linguistic expressions that reflects an established conceptual metaphor. Here, in two empirical studies, we investigated this linguistic innovation. In particular, we established whether left-right linguistic metaphors are conventional for members of the US military, compared to civilians, and explored the subtle conventions that govern the use of these expressions.

7.3 Study 10 Methods

In Study 10, participants rated the acceptability of sentences. Features of the sentences allowed us to measure whether, when, and to whom lateral (left-right) linguistic metaphors are acceptable.

7.3.1 Participants

Active Duty members of the US military (n = 23) participated for $10, and civilian undergraduates at UC San Diego (n = 31) participated for course credit. The military
participants included 4 Army, 1 Navy, and 18 Air Force. They included 8 Officers (leaders who hold positions of authority) and 15 Enlisted (non-Officer) members. We recruited in person on a joint military post, as well as through word of mouth in the US military community. All participants were naive to the goals of the study. Data collection was planned to continue until we reached 25 military participants or the end of the academic quarter. Because civilian data collection was easier, we aimed for a slightly larger civilian sample.

7.3.2 Materials

We included the following sentence types:

A. **Standard**, which used *earlier* or *later* to describe the rescheduling of an event (e.g., *The meeting was moved two days earlier, from Friday to Wednesday.*);

B. **Dynamic-Lateral**, which used *left* and *right* to describe the rescheduling (e.g., *The meeting was moved two days to the left, from Friday to Wednesday.*);

C. **Static-Lateral**, which used *left* and *right* to describe the relative timing of an event which was not rescheduled (e.g., *The meeting on Wednesday is two days to the left of the meeting on Friday.*);

D. **Ungrammatical**, which used the same words as a Standard sentence but in a scrambled order (e.g., *From the meeting was two earlier days, Friday to Wednesday moved.*).

To examine the generality of any eventual linguistic conventions, we varied these sentences along a number of dimensions: whether the event occurred earlier or later; whether the description used the timescale of hours, days, or months; and whether the description crossed a standard temporal boundary or remained entirely within a standard period of time (e.g., an event is rescheduled to later in the same week, from Tuesday to Friday, or to another week entirely, from Friday to the following Monday.) These were
fully crossed within each sentence type, for a total of 12 sentences of each type.

7.3.3 Procedure

The study was completed on a computer. Participants were instructed to imagine a new colleague whose native language was not English and rate the acceptability of sentences (n = 48) uttered by this colleague, based on how participants would normally talk at work. Acceptability ratings used a 7 point Likert scale (1=totally unacceptable, 7=totally acceptable).

All participants saw the same 48 sentences (12 per phrase type) in a random order. Each sentence was presented on its own page. Participants then supplied standard demographic information (education, age), and military participants reported their service branch (Army, Navy, etc.), rank, and the year they joined the service. No other measures were collected.

7.3.4 Exclusions and Analyses

We subtracted each person’s mean Ungrammatical rating from their mean Standard rating. Three participants (1 military, 2 civilians) did not rate Standard sentences at least one point higher than Ungrammatical ones, and were eliminated from further analysis. Ratings were standardized by participant (i.e., z-scored), and then analyzed using linear mixed-effects models. Models used the maximal converging random effects structure justified by the experimental design [6], with random intercepts and slopes for both participants and items. For analyses that include the 4 sentence conditions, conditions were forward-difference coded: Standard > Dynamic-Lateral > Static-Lateral > Ungrammatical. Analyses that distinguished between Officers and Enlisted personnel were also forward-difference coded: Civilian > Enlisted > Officer.
7.4 Study 10 Results

Ratings did not differ by timescale (i.e., hours, days, and months; p = .80) or by boundary (i.e., whether the rescheduled meeting was in the same or a new day, week, or year), so we collapsed timescales and boundaries for all subsequent analyses.

We first verified that participants from both populations rated the Standard phrases as most acceptable and the Ungrammatical phrases the least acceptable, with the Lateral phrases in between. Standard items were rated as more acceptable (M = 0.96) than Dynamic-Lateral phrases (M = 0.21, b = 0.75 0.14 SEM, t = 5.5, p < .001), which were more acceptable than Static-Lateral phrases (M = -0.23, b = 0.44 0.10 SEM, t = 4.6, p < .001), which in turn were more acceptable than Ungrammatical phrases (M = -0.94, b = 0.71 0.13 SEM, t = 5.3, p < .001). As predicted, there was no evidence that military and civilian participants differed in their ratings of Standard phrases (M_{civilian} = 1.02, M_{military} = 0.91; b = 0.11 0.12 SEM, t = 0.9, p = .38), though they did differ in their relative unacceptability of Ungrammatical phrases, with military members especially disapproving of this type of phrase (M_{civilian} = -0.75, M_{military} = -1.10; b = 0.35 0.14 SEM, t = 2.5, p = .02).

We next tested our critical prediction: That these patterns of acceptability for lateral phrases would differ systematically for civilians and members of the military. Indeed, we found an interaction between phrase condition (Standard vs. Dynamic-Lateral) and population (b = 0.49 0.23 SEM, t = 2.2, p = .04), demonstrating that while the two groups did not differ in the acceptance of Standard phrases, military members were more accepting of Dynamic-Lateral phrases than civilians were. There was also an interaction between Dynamic-Lateral and Static-Lateral phrases and the two populations (b = -0.31 0.14 SEM, t = -2.3, p = .03): while military members accepted Dynamic-Lateral phrases more than civilians did (M_{military} = 0.39, M_{civilian} = 0.00), the two groups converged on
acceptability judgments of Static-Lateral phrases ($M_{\text{military}} = -0.20, M_{\text{civilian}} = -0.27$).

To investigate whether these lateral expressions were more acceptability among those with more military responsibility – e.g., among Officers compared to Enlisted personnel – we added a fixed effect of subpopulation (civilian = 1, Enlisted = 2, Officer = 3; centered so the mean was 0). See Figure 7.1. We replicated the main effects that Standard sentences were more acceptable than Dynamic-Lateral ($b = 1.01 \pm 0.18$ SEM, $t = 5.7, p < .00014$), which were in turn more acceptable than Static-Lateral ($b = 0.03 \pm 0.12$ SEM, $t = 2.4, p = .02$), which again were more acceptable than Ungrammatical ($b = 0.49 \pm 0.12$ SEM, $t = 2.7, p = .009$). These selective preferences, however, differed by military subpopulation: compared to Enlisted members, Officers rated Dynamic-Lateral (vs. Standard) phrases as more acceptable ($b = -0.65 \pm 0.25$ SEM, $t = -2.5, p = .02$). As a result of Officers’ increased acceptance of Dynamic-Lateral phrases, the difference in acceptability for Dynamic-Lateral compared to Static-Lateral phrases was greater for Officers than Enlisted members ($b = 0.44 \pm 0.15$ SEM, $t = 3.0, p = .005$). Officers were also marginally less accepting of Ungrammatical phrases (compared to Static-Lateral) than Enlisted members were ($b = 0.50 \pm 0.26$ SEM, $t = 2.0, p = .06$). There were no other interactions between phrase type and subpopulation ($bs < 0.26, ps > .40$). These results indicate that Officers, whose roles tend to include the most leadership and planning, are especially accepting of Dynamic-Lateral sentences, as compared not only to civilians, but to their Enlisted counterparts as well.

Finally, as an exploratory analysis of the nuances of this emerging linguistic convention, we investigated whether meetings that were moved later (to the right) differed in acceptability from those moved earlier (to the left), as suggested by initial anecdotal evidence. Indeed, across both military and civilian populations and all phrase types, meetings that were postponed were rated as more acceptable than those rescheduled to an earlier time ($b = 0.22 \pm 0.09$ SEM, $t = 2.4, p = .02$), perhaps reflecting the relative frequency
Figure 7.1: Results of Study 10: Compared to civilians, military personnel (especially Officers) were more accepting of Dynamic-Lateral phrases (e.g., meeting was moved to the left). There was no difference in civilians and military members acceptance of Static-Lateral phrases. Error bars = SEM.

of rescheduling meetings to later times compared to earlier times in real life. Further, we found an interaction between phrase type (Dynamic-Lateral vs. Standard), population (military vs. civilian), and reschedule direction (b = -0.40 0.15 SEM, t = -2.7, p = .006). We thus zoomed in on these Dynamic-Lateral phrases, while distinguishing between Officer and Enlisted members of the military. This analysis revealed that, compared to civilians, both Officer and Enlisted members especially preferred phrases in which a meeting was moved to the right over those in which a meeting was moved to the left (Enlisted: b = -0.25 0.12, t = -2.09, p = .04; Officer: b = 0.23 0.10 SEM, t = 2.3, p = .02). Officer and Enlisted members did not differ significantly from each other (b = 0.02 0.13, t = 0.14, p = .89). Therefore, the acceptability of dynamic-lateral phrases among members of the military appears to be especially pronounced for particular usages (i.e., when events are moved later).

Finally, we verified that the main effects were not driven by any demographic differences between the three populations. While Officers and Enlisted members did not differ in age (t(12.7) = 0.05, p = .96) or years of military experience (t(8.2) = 0.2, p = .81),
they were significantly older than the civilian participants ($t(22.8) = 8.4, p < .00001$).

Nevertheless, accounting for age did not affect any of the critical findings. We added the main effect of participant age as a fixed effect and repeated the most specific mixed effect model, which included sentence condition (forward-difference coded so that Standard > Dynamic-Lateral > Static-Lateral > Ungrammatical) and subpopulation (forward-difference coded so that civilian > Enlisted > Officer). As in the analysis that did not account for participant age, Standard sentences remained more acceptable than Dynamic-Lateral ($b = 1.01 \pm 1.8$ SEM, $t = 5.7, p < .0002$), which were in turn more acceptable than Static-Lateral ($b = 0.27 \pm 0.12$ SEM, $t = 2.4, p = .02$), which were more acceptable than Ungrammatical ($b = 0.49 \pm 0.18$ SEM, $t = 2.7, p = .009$).

### 7.5 Study 10 Discussion

This study confirmed that a system of lateral metaphors has become conventionalized among members of the US military, a subculture of American English speakers. If this reflected a general conceptual or linguistic difference – perhaps a willingness among military personnel to think about time along a left-to-right timeline – then this should have been reflected in increased acceptability for all lateral expressions. Instead, military personnel were especially accepting of lateral metaphors that used the dynamic language of movement, especially when used to delay an event, suggesting that this is a targeted linguistic convention, restricted to particular linguistic constructions.
7.6 Study 11: Lateral Linguistic Metaphors in Different Contexts

Study 11 was designed to replicate and further explore this linguistic conventionalization. In addition, we sought to determine whether military personnel are aware that lateral metaphors are specific to varieties of English spoken in the military, and not shared with the larger civilian population.

We also aimed to confirm that differences between military and civilian acceptability ratings truly reflected linguistic differences – rather than differences in familiarity or expertise with a lateral construal of time, which could affect the phrases’ interpretability. To rule out the possibility that military members were just more skilled or familiar with a lateral, left-to-right timeline, we also included a forced-choice interpretation task, where participants had to indicate when an event would occur if it were moved to the left or to the right.

7.7 Study 11 Methods

7.7.1 Participants

Members of the US military (n = 29) participated for $10. Civilian undergraduates (n = 36) participated for course credit. Military participants included 6 Army, 6 Navy, 9 Air Force, and 3 Marines (5 did not identify their branch); 10 were Officers, 12 were Enlisted, and 7 did not identify their rank. Of these, 11 were veterans (often still working alongside active duty military members), and 18 were active duty. Data collection was planned to continue until we reached 36 military participants or the end of the academic quarter.
7.7.2 Materials

Materials were identical to Study 10 with two differences. First, to confirm that lateral metaphors can be used with a variety of verbs, we also included items that used the verb *push* (e.g., *pushed* two months), in addition to items using the verb *move* as in Study 10. Second, to reduce the total number of items, we did not vary the timescale (i.e., weeks, months, or years), since it had no effect on acceptability in Study 10.

7.7.3 Procedure

Participants completed two randomly-ordered tasks: the Acceptability Rating task from Study 10, and a forced choice Sentence Completion task.

The Acceptability Rating task was based on Study 10, with one critical difference: participants completed two randomly-ordered blocks of acceptability ratings, one in which they were asked to imagine all their colleagues were in the military, and another in which they imagined all of their colleagues were civilians. Manipulating the utterances context in this way allowed us to test whether military participants were sensitive to the community-specificity of the lateral-dynamic metaphors.

In the forced-choice Sentence Completion task, participants read the same sentences as in the Acceptability Rating task, but with a blank in place of month (e.g., *The meeting was moved two months to the right, from November to ___*) Choices included all odd-numbered months (January, March, etc.) and *I don't know*.

To refresh participants between these tasks, they completed a brief “spot the differences” game, in which they had 45 seconds to count as many small differences as possible between two nearly identical images.

Participants were randomly assigned to complete the acceptability ratings (both in a military and civilian context) or forced choice block first. For the acceptability ratings,
context order (i.e., military vs. civilian context) was assigned randomly. There were 32 sentences for each of the three task components; every participant saw every sentence in a random order.

### 7.7.4 Exclusions & Analyses

Exclusions and Analyses were unchanged from Study 10. We first investigated main effects and interactions by generalizing participants as either military members or civilians. When members of these two populations differed in their responses, we further divided the military members by whether they were Officers or Enlisted personnel, as in Study 10. Participants were eliminated from further analysis if their mean rating for Ungrammatical sentences was not at least one point less than their mean rating for Standard sentences. This excluded five participants (2 military, 3 civilians). Veteran status did not affect judgments (ps > .26), so analyses collapsed across current and former military personnel. Similarly, there were no effects of verb (move vs. push), so analyses were collapsed over both verbs.

### 7.8 Study 11 Results

#### 7.8.1 Acceptability Ratings

Do military members and civilians view the different phrase types?

As in Study 10, we first verified that participants from both populations rated the Standard phrases as most acceptable and the Ungrammatical phrases the least acceptable, with the Lateral phrases in between. Once again, Standard items were rated as more acceptable than Dynamic-Lateral phrases (M = 0.99 vs. M = 0.25; b = 0.71 ± 0.09 SEM, t = 7.5, p < .0001), which were more acceptable than Static-Lateral phrases (M = -0.03;
b = 0.28 0.06 SEM, t = 4.7, p < .0001), which in turn were more acceptable than Ungrammatical phrases (M = -1.21; b = 1.2 0.10 SEM, t = 11.9, p < .0001).

We next attempted to replicate our main finding from Study 10: That, compared to civilians, military participants had a selective preference for lateral metaphors, especially dynamic ones. We replicated our critical finding that the populations differed in their preference for Dynamic-Lateral phrases compared to Standard ones (b = 0.74 0.18 SEM, t = 4.1, p < .001). Moreover, acceptability for Static-Lateral (vs-Dynamic-Lateral) phrases differed between the two populations (b = -0.21 0.18 SEM, t = 4.1, p = .0001): military personnel reported a difference in acceptability between Static-Lateral and Dynamic-Lateral phrases that was more than double that of civilians (military: Dynamic-Lateral M = 0.45, Static-Lateral M = 0.06; civilians: Dynamic-Lateral M = 0.08, Static-Lateral M = -0.10), suggesting a strong, selective preference by military personnel for dynamic, rather than static, lateral metaphors. In order to investigate whether the populations differed in their overall preferences for Static-Lateral phrases (i.e., compared to Standard phrases), we constructed another model in which Standard phrases were treated as a baseline against which all other phrase types were compared. This revealed that, overall, military members thought that Static-Lateral phrases were more acceptable than civilians did (b = -0.99 0.10 SEM, t = 9.8, p < .00001), replicating a numerical but statistically non-significant effect from Study 10. Thus, all forms of lateral metaphors were rated as more acceptable by military members, compared to civilians, although this effect was especially pronounced for dynamic, rather than static, phrases.

As in Study 10, we next investigated whether this linguistic convention was especially entrenched among military Officers, and present but less pronounced among Enlisted military personnel. Civilians and Enlisted personnel differed in their acceptance of Dynamic-Lateral phrases, compared to Standard ones (b = 0.54 0.24 SEM, t = 2.3, p = .03). While both civilians and Enlisted personnel thought Dynamic-Lateral phrases were
less acceptable than Standard ones, this dispreference was twice as large for civilians
(b = 1.07  0.12 SEM, t = 8.9, p < .0001) than for Enlisted personnel (b = 0.55  0.24
SEM, t = 2.2, p = .04). By contrast, both civilians and Enlisted military personnel
thought that Static-Lateral phrases were worse than Dynamic-Lateral ones (for civilians:
b = 0.18  0.05, t = 3.3, p = .003; for Enlisted: b = 0.35  0.12 SEM, t = 2.9, p = .01),
and this dispreference did not differ by population (b = -0.18  0.13 SEM, t = -1.4, p =
.16). Enlisted military personnel, therefore, had a systematic and targeted preference for
Dynamic-Lateral sentences, compared to civilians.

We next examined whether these preferences were even more pronounced among
Officers. Compared to Standard phrases, Officers rated Dynamic-Lateral phrases as more
acceptable than Enlisted personnel did (b = -0.65  0.25 SEM, t = 2.5, p = .02). To explore
this effect, we examined the main effect of phrase type (Dynamic-Lateral vs. Standard)
in Officers and Civilians separately. While both Officers and Civilians gave numerically
lower ratings for Dynamic-Lateral phrases compared to Standard phrases, this difference
was small and non-significant for Officers (b = 0.23  0.21, t = 1.1, p = .30), but five-times
larger and highly significant among Civilians (b = 1.07  0.12 SEM, t = 8.9, p < .0001).
Even compared to their Enlisted personnel, military Officers are especially accepting
of Dynamic-Lateral phrases not even distinguishing them, statistically, from Standard
phrases.

In addition, unlike Civilians and Enlisted members, Officers did not make a
strong distinction between Dynamic-Lateral and Static-Lateral phrases: Officers thought
Static-Lateral phrases were only slightly and non-significantly less acceptable than
Dynamic-Lateral ones (M_{Static} = 0.16 vs. M_{Dynamic} = 0.49, b = 0.33  0.20 SEM, t = 1.7,
p = .13). To test whether Static-Lateral phrases were considered worse than Standard
phrases, we recoded phrase type so Standard was the baseline against which all other
levels were compared. This confirmed that, among Officers, Static-Lateral sentences
were only marginally less acceptable than Standard ones \( (b = 0.56 \ 0.28 \text{ SEM}, t = 2.0, p = .07) \). Thus, for Officers, Dynamic-Lateral phrases were no worse than Standard phrases, and Static-Lateral phrases were only marginally worse than Standard ones.

Finally, we examined whether acceptability judgments differed based on whether a meeting was rescheduled to a later time (to the right) or to an earlier one. As in Study 10, meetings moved later were more acceptable than those moved earlier \( (M_{\text{Later}} = 0.05 \) vs \( M_{\text{Earlier}} = -0.05, b = -0.09 \ 0.02 \text{ SEM}, t = -4.0, p = .0002) \). This finding was marginally stronger for Standard sentences (those that used the term earlier or later) than Dynamic-Lateral (using the term left or right; \( b = -0.01 \ 0.06 \text{ SEM}, t = -1.83, p = .07) \). However, military and civilians did not differ in their preference for rescheduling meetings to later times over earlier ones \( (b = 0.02 \ 0.03, t = 0.74, p = .46) \).

7.8.2 Does context affect phrases’ acceptability?

Next we investigated whether an utterance’s context – whether it was uttered in a setting with civilians or members of the military – had an effect on its acceptability (Figure 7.2 & Table 7.1). A model that added an additional fixed effect for context replicated the central findings reported above. Standard sentences were more acceptable than Dynamic-Lateral ones, which were in turn more acceptable than Static-Lateral, which were finally more acceptable than Ungrammatical \( (bs > .29, ps < .00001) \). And, once again, there was a selective acceptance of Dynamic-Lateral phrases by members of the military, compared to civilians (civilian vs. Enlisted; \( b = 0.54 \ 0.24 \text{ SEM}, t = 2.3, p = .03 \); enlisted vs. Officer: \( b = 0.31 \ 0.30 \text{ SEM}, t = 1.0, p = .31 \)).

Overall, there was an effect of context \( (b = -0.14 \ 0.2 \text{ SEM}, t = -8.0, p < .00001) \), with higher acceptability for phrases in a military than civilian work setting (unstandardized ratings: \( M = 3.40 \) vs. \( 3.22 \)). Standard phrases were seen as marginally more acceptable in a military than a civilian setting (unstandardized ratings: \( M = 5.41 \)).
Civilians and Enlisted members did not differ in their acceptance of Standard phrases across contexts (b = 0.01, p = .90), but Officers rated such sentences as marginally more acceptable in a military than a civilian context (compared to Enlisted, b = 0.20 0.10 SEM, t = 1.9, p = .05).

We next zoomed in on the phrases of especial interest, the Dynamic-Lateral ones. This subset analysis revealed that acceptability differed across contexts, (b = -0.26 0.03 SEM, t = -7.4, p < .0001), with people viewing them as significantly more acceptable in a military than a civilian context – more than 4 times more so than for Standard phrases (unstandardized ratings: M = 3.99 vs. M = 3.60). This main effect of context was driven by an interaction with population (b = -0.49 0.06, t = 8.3, p < .0001). While civilians did rate the Dynamic-Lateral phrases (compared to Standard) as more appropriate in a military than a civilian context (b = 0.04 0.02 SEM, t = 2.2, p = .03), military members demonstrated context-selective acceptability for these phrases to a much greater extent (almost 6 times more; b = -0.22 0.02 SEM, t = -9.5, p < .0001). Follow-up analyses of military members revealed this context-sensitivity was present among Enlisted personnel, who rated Dynamic-Lateral phrases as more acceptable in a military than a civilian context compared to civilians (b = 0.26 0.08 SEM, t = 3.4, p = .0005) but even more pronounced for Officers, for whom the effect was nearly 50% greater than it was for Enlisted military personnel (b = 0.39 0.10, t = 4.1, p < .0001).

When we analyzed Static-Lateral phrases on their own, we found that these phrases were also rated as more acceptable in military workplaces than in civilian ones (b = -0.15 0.03 SEM, t = -4.6, p < .0001). Once again, this was driven by an interaction between context and population (b = 0.24 0.06, t = 4.1, p < .0001), with military members selectively preferring Static-Lateral phrases in a military context. We again further analyzed this data using the Officer-Enlisted distinction for military members. The difference in acceptability for Static-Lateral phrases across contexts was greater for
Enlisted members than civilians (b = 0.28 0.07 SEM, t = 3.8, p = .0001), but not for Officers compared to Enlisted (b = -0.08 0.09 SEM, t = -0.9, p = .37). However, when compared to Dynamic-Lateral phrases, civilians did not rate Static-Lateral sentences as less appropriate in a military than civilian context (b = 0.04 0.06 SEM, t = 0.7, p = .48), nor did Enlisted members (b = 0.06 0.10 SEM, t = 0.6, p = .58), but Officers did (b = -0.42 0.08 SEM, t = -5.1, p < .0001).

Thus, military personnel appear to recognize that their use of lateral metaphors is restricted to their variety of English, and would be unacceptable in contexts that were populated primarily by civilians. This metalinguistic awareness was especially pronounced among officers.

Figure 7.2: Results of Study 11: Effect of context and phrase type on acceptability. Overall, military personnel were more accepting of lateral metaphors, in both static and dynamic forms, although this was restricted to contexts where interlocutors were members of the military (right) rather than civilians (left). (Error bars = SEM.)

7.8.3 Forced Choice

One alternative explanation for these results is that, rather than reflecting a novel linguistic convention, these differences in acceptability are merely a consequence of increased familiarity among military personnel with nonlinguistic uses of lateral space
Table 7.1: Z-scored acceptability ratings for participants by sentence type, military status and rank, and context.

<table>
<thead>
<tr>
<th>Sub-population</th>
<th>Civilians</th>
<th>Enlisted</th>
<th>Officers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Standard</td>
<td>Dynamic-Lateral</td>
<td>Static-Lateral</td>
</tr>
<tr>
<td></td>
<td>Civilian</td>
<td>Military</td>
<td>Civilian</td>
</tr>
<tr>
<td>Standard</td>
<td>1.16</td>
<td>1.16</td>
<td>0.86</td>
</tr>
<tr>
<td>Dynamic-Lateral</td>
<td>0.11</td>
<td>0.06</td>
<td>0.22</td>
</tr>
<tr>
<td>Static-Lateral</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.17</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>-1.08</td>
<td>-1.21</td>
<td>-1.23</td>
</tr>
</tbody>
</table>

To represent time. We addressed this concern with the forced choice interpretation task, which investigated whether individuals could interpret these novel expressions, even if they judged them to be linguistically unacceptable. These results are shown in Table 7.2.

Table 7.2: Performance on the forced choice interpretation task, for each of the phrase types and for both populations. Mean accuracy (SD)

<table>
<thead>
<tr>
<th></th>
<th>Civilians</th>
<th>Military</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>96.6 (8.4)</td>
<td>99.1 (3.3)</td>
</tr>
<tr>
<td>Dynamic-Lateral</td>
<td>93.9 (18.3)</td>
<td>95.4 (7.1)</td>
</tr>
<tr>
<td>Static-Lateral</td>
<td>73.9 (31.7)</td>
<td>93.9 (21.6)</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>78.8 (29.1)</td>
<td>78.2 (37.9)</td>
</tr>
</tbody>
</table>

We first analyzed accuracies for each sentence type by population. There was no difference in overall accuracy between civilians and military members (p = .19). Participants were overall less accurate for Static-Lateral (78.9%) than Dynamic-Lateral statements (94.7%; b = 2.07 0.63 SEM, z = 3.3, p = .001; Table 2), but this did not differ by population (military vs. civilian; p = .30). There were no other main effects of phrase type or interactions between phrase type and population (ps > .15).

Thus, military and civilian participants showed no differences in the ability to interpret the sentences with lateral metaphors, despite differing in how acceptable they found them. This suggests that the increased acceptance of lateral phrases among
members of the military did not merely reflect improved interpretability, but instead reflects the emergence of a genuine linguistic convention.

7.9 Discussion

Can shared ways of of thinking metaphorically give rise to new conventionalized ways of of speaking metaphorically? In two studies, we found that members of the U.S. Military, a subcommunity of American English speakers, have appropriated and conventionalized a lateral metaphor for time. While English speakers generally conceptualize time as a left-to-right path, members of the US military have begun to accept metaphorical expressions like *move the meeting to the right* that use the language of lateral space to describe temporal relations. This case study fills an evidential gap, since *thought-first* accounts assert that systematic metaphorical expressions are the linguistic manifestation of pre-existing systems in thought (e.g., [74]). Demonstrating this purportedly pervasive relationship has been challenging, since, in linguistic communities around the world, conceptual metaphors typically already have systematic linguistic counterparts. Thus, the adoption of lateral spatial metaphors by the U.S. Military provides a unique and rare opportunity to document the shift from a cultural pattern of thought to a conventionalized system of linguistic metaphors.

The current results contribute to our understanding of the causal relations between metaphorical thought and language. Learning to use novel linguistic metaphors is known to create new conceptual metaphors in the speaker (Experiment 1 [65]). In one study, English-speaking participants learned linguistic metaphors that placed earlier events either above or below later ones (i.e., *breakfast is above dinner* or *breakfast is below dinner* [65]); afterward, they exhibited metaphor-consistent responses on an implicit measure of their mental space-time associations. Our results suggest that the causal
arrow can also point in the opposite direction. Not only can metaphorical language shape mental representations, but our mental representations can also make their way into conventionalized language. There appear to be bidirectional causal relationships between metaphorical language and metaphorical thought.

We also aimed to distinguish whether the documented practice was solely a case of lexical substitution (i.e., regardless of surrounding language context, people substitute *right* for *later* and *left* for *earlier*) or marked the emergence of an entire system of linguistic conventions. While systems of linguistic metaphor are typically productive, allowing speakers to generate novel expressions, they also tend to be governed by idiosyncratic conventions. For instance, in English one can say that a meeting was *pushed up* to an earlier time, but not that a meeting was *pushed down* to a later time. In the case of lateral metaphors, military personnel preferred expressions that involved moving an event (i.e., Dynamic-Lateral phrases) over similar phrases that merely described temporal relations (i.e., Static-Lateral phrases). There were also hints that these lateral phrases were most acceptable when moving an event later (to the right). These quirks suggest the emergence of idiosyncratic conventions governing the use of this metaphor system. On the other hand, lateral expressions were acceptable with different verbs *push* and *move* which suggests that the metaphor affords a certain degree of creative productivity. These lateral expressions, therefore, have many of the hallmarks of a conventionalized system of linguistic metaphor, exhibiting both idiosyncratic constraints and creative productivity.

**The birth and life of lateral metaphors for time**

Why have military members, in particular, adopted lateral metaphors that are absent from civilians’ language? This answer will likely involve factors operating at different levels, from the individual to the institutional, all working together to support and regiment the use of this linguistic metaphor system. At the smallest unit of analysis,
there may be something special about people who are inclined to join the military. Perhaps they are predisposed, for some reason, to adopting lateral linguistic metaphors for talking about time. However, this seems unlikely – and even if it were true, it would probably be insufficient to account for the widespread establishment of a novel linguistic convention. Instead, we propose that a suite of specific cultural and socio-institutional features of the US military created an optimal environment for the emergence of a system of lateral metaphors, which prior work has suggested supports the emergence of conceptual metaphors [50].

At the cultural level, the initial emergence and widespread acceptance of these linguistic metaphors may have been encouraged by the ubiquitous and highly regimented use of a specific kind of temporal artifact. Duty Rosters are documents that keep track of the assignments for each member of a military unit, and are thus a critical and recurring part of military work-life. Duty Rosters are standardized and governed by specific regulations (Army Regulation [AR] 220-45). Each row represents an individual. Each column represents a successive date, ordered from left to right. Unlike standard American calendars, in which each row only has 7 days across, Duty Rosters arrange days in a continuous line extending rightward potentially endlessly, when implemented as a digital spreadsheet. These artifacts could have created recurring contexts in which left-right language became both natural and efficient.

In addition, the culture of the US military puts a high value on minimizing ambiguity – as demonstrated, for example, by their use of a 24-hour clock instead of the more common American practice of using a 12-hour clock. Unlike many of the standard phrases that currently exist in Standard English for speaking about temporal sequences, lateral metaphors are completely unambiguous with regards to the direction of temporal change [38]. When a meeting is moved forward, it could have been moved earlier or later, depending on ones construal [9]. But when a meeting is moved to the left, there is
only one possibility it was moved earlier. Given their enculturation to avoid ambiguity, members of the US military may have been inclined to accept and conventionalize these lateral metaphors.

At the socio-institutional level, the US military is both highly insular and hierarchical. Military members often live and work together in closed communities – and when they are working internationally, they can be surrounded by locals with whom they do not share a language. Within such closed communities, linguistic innovations may be more likely to spread, and new members may be incentivized to adopt non-standard linguistic conventions as a way to achieve in-group status. Moreover, the military is marked by a highly hierarchical, top-down power structure. Thus, conventions, linguistic or otherwise, that originate at the top of the institution are likely to percolate down throughout the rest of the military – for instance, from officers who adopted lateral metaphors to talk about Duty Rosters, to the enlisted members who must carry out their assigned roles. By combining insularity, a strictly hierarchical power structure, and highly regimented cultural artifacts that make extended use of lateral timelines, the US military may have created a perfect environment for the emergence and establishment of this novel system of linguistic metaphors.

Another remaining question is whether and how this linguistic innovation may spread outside the US military to the larger community of English speakers. While the military community is quite insular, it is not completely disconnected from the civilian world. Military officers, in particular, often transition to civilian careers – in part because, at the higher levels of the military, there are fewer opportunities for promotion than there are officers vying for them. These officers, as we have shown, also happen to be the segment of the US military who have most reliably adopted this new metaphor system. Military officers are thus a likely vector for the spread of this metaphor system into the larger community of English speakers.
Moreover, as demonstrated by their success on the interpretation task (Study 11), civilians can make sense of these lateral metaphorical expressions, however unusual they may seem. This confirms a major prediction of thought-first accounts of the origin of systematic linguistic metaphors: Our civilian participants could make sense of these metaphors, presumably, because they already have an analogous conceptual metaphor [120, 43, 93, 128]. It is thus a small leap for civilians to adopt these systematic linguistic expressions as a conventionalized part of their language. In addition, as discussed above, these expressions offer the benefit of reducing ambiguity, compared to equivalent phrases in Standard English (e.g., moved ahead can mean earlier or later). We predict that civilians will eventually adopt this system of lateral metaphors, driven by their alignment with a preexisting conceptual metaphor, the mobility of military officers, and the expressions’ lack of ambiguity.

**Future Work & Conclusions**

This work opens the question of whether there are cognitive consequences of adopting lateral metaphors for time. Can adopting lateral linguistic metaphors facilitate reasoning about temporal change? Does it reduce miscommunication (e.g., allowing speakers to avoid ambiguous descriptions like Wednesdays meeting was moved forward two days)? Or might it increase other kinds of miscommunication, for example when English speakers communicate with Hebrew and Arabic speakers, whose mental timelines run right to left, counter to English speakers [120]?

Together, the two studies presented here provide a proof-of-concept of an oft-theorized causal influence of metaphorical thought on language: not only can linguistic metaphors give rise to new ways of thinking, but entrenched patterns of thinking can also give rise to new metaphors in language. Given the right cultural milieu, metaphors can move from mind to mouth.
Chapter 7, in part, has been submitted for publication of the material as it may appear in Hendricks, R. K., Bergen, B. K. & Marghetis, T. (2017). Do metaphors move from mind to mouth? Evidence from a new system of linguistic metaphors for time. *Cognitive Science*. The dissertation author was the primary investigator and author of this paper.
Chapter 8

Conclusions

How do we construct our knowledge of abstract concepts we can’t physically experience? To what extent can language, and in particular, linguistic metaphor, shape our mental representations? These are core questions that sit at the intersection of two prominent lines of work in Cognitive Science – one focused specifically on metaphor, and the other on the relationship between language and thought. The experiments included in this dissertation bridge these two lines of work to advance our understanding of the role that linguistic metaphor plays in how we construct mental representations of abstract concepts, using the domain of time as a test case.

Although extensive work has shown cross-linguistic differences in thought that correspond to differences in language (in the domain of time, e.g., [94, 10, 13, 44, 87, 58, 73, 131]), most work on this topic has been quasi-experimental. Since participants cannot be randomly assigned to be a native speaker of a particular language, prior studies have not been able to rule out the numerous confounds inherent when comparing performance by speakers of different languages.

Using this limitation in prior research as a starting point, I opened this line of work by asking: can learning a new system of linguistic metaphors shape the way people
think about time? And if so, can language create nonlinguistic mental representations? To this end, in Experiment 1 English speakers learned new metaphor systems, which referred to earlier events as either above or below later ones, and then completed a nonlinguistic space-time association task while undergoing either verbal or spatial interference. Participants showed space-time associations that were congruent with their newly-learned metaphor systems, suggesting that learning a new way to talk can create new ways of thinking. Further, those associations were unaffected by verbal (or spatial) interference, lending support to the idea that patterns learned through language can create nonlinguistic representations. Even if a representation originated in language, it is not necessarily linguistic in our minds.

Experiment 2 showed that the verbal interference task used in Experiment 1 (compared to the spatial interference task used) does alter performance on another task that is known to invoke a linguistic routine – counting objects. This finding thus strengthened our confidence in the inference that newly-learned representations of time originating in language can create nonlinguistic patterns of thought.

The following two experiments increased the ecological validity of the findings from Experiment 1. These experiments focused on Chinese-English bilinguals, who have experience with natural language metaphors in Chinese that use vertical space to talk about time. The participants completed the same nonlinguistic space-time association task under the same interference types as in Experiment 1, without any linguistic intervention in the lab (Experiment 3) or after learning metaphors in English that referred to earlier events as either above or below later ones (Experiment 4; same metaphor training procedure as in Experiment 1). Together, these experiments showed that vertical mental representations of time consistent with natural language metaphors are comparable in magnitude and nature (unaffected by interference) to those newly learned in the lab.

The next group of experiments applied the same methods to representations
on the lateral axis, first examining lateral space-time associations (Experiment 5, and Experiments 6 & 7 with greater delays between encoding and recalling interference), and then examining lateral space-number associations (Experiments 6 & 7). These experiments provided an opportunity to examine spatial representations arising from different sources – since we don’t typically talk about time by referring to lateral terms (like left or right), these representations are more likely results of cultural practices like reading, writing, and artifact use. Those experiments revealed that space-time and space-number congruency effects were unaffected by interference, pointing to a commonality in the nature of spatial representations for the concepts of time and number, even when they have arisen from different origins.

Experiment 8 ruled out the possibility that the spatial dual task hadn’t taxed working memory for the relevant feature of space as an explanation of the representations’ persistence under this interference type. In that experiment, participants again made time judgments by pressing response keys that were arranged laterally while undergoing interference, but this time all participants underwent spatial interference – the key manipulation was whether the interference task was oriented vertically (perpendicular to the relevant axis for the task) or horizontally (consistent with the relevant axis for the task.). This manipulation did not alter the effect of spatial interference on space-time associations, suggesting that either a spatial dual task is an inadequate method for probing the necessity of spatial working memory capacity in the moment when space-time associations manifest, or that these cognitive resources may not actually need to be recruited for us to observe systematic space-time associations.

Experiment 9 built on the same line of work by examining another kind of nonlinguistic manifestation of representations arising from linguistic metaphor – namely, whether such representations can be seen in the body through co-speech gesture. As in prior work, English-speaking participants demonstrated canonical left-to-right timelines
in their gestures, placing earlier events to the left of to later ones. Their gestures also reflected their newly-learned metaphors by exhibiting a vertical component that tended to be consistent with their newly-learned linguistic metaphors. Thus, this experiment suggested that representations shaped by metaphors in language can spill out into the body.

The final two experiments turned the tables to examine the opposite causal relationship; that patterns in thought shape language. Many accounts advocate for a thought-first perspective, in which systematic linguistic metaphors arose from pre-existing patterns in thought [74, 112, 49, 37]. However, it has been challenging to find direct evidence of this process, as a result of the slow pace of language change and the pervasiveness of metaphors that are already established in both thought and language. Studies 10 & 11 document a sub-group of English speakers who have begun to use such a system of linguistic metaphors: members of the U.S. Military. This community of English speakers systematically uses the language of lateral space (left, right) to describe temporal order (earlier, later). This case study informs theory about the origins of linguistic metaphor, by demonstrating that shared ways of thinking metaphorically can become crystallized in language.

**Overarching Conclusions**

This line of work advances our understanding of the relationship between metaphor and thought. It sheds light on the bidirectional nature of mental representations and linguistic metaphors. That is, systems of linguistic metaphor can create nonlinguistic mental representations, and cultural patterns in thought can give rise to systems of linguistic metaphor. Together, these findings inform why we have so many systems of metaphors that do not include just a single template expression used space to talk about concepts like time, but instead apply associations like those between space and time broadly and
productively *(approaching* deadlines, with weekends *coming up*, reflecting *back* on the past, and planning *ahead*, for example), and why extensive research has revealed that these space-time correspondences in language are also evident in thought.

One theme evident throughout these experiments is the dynamic nature of our mental representations of time and other abstract concepts. These representations can be shaped by a myriad of influences, including gesture [83], reading and writing experience [120, 43, 8, 15], use of artifacts [4], cultural priorities [32], and, as shown here, by metaphors in language.

Further, such representations can be shaped by short-term influences, like the approximately 10-minute metaphor training session employed in Experiments 1, 4, and 9 of this work. Evidence of these representations can also change over the course of an experiment block as “counter-trainings” are introduced, which was demonstrated by congruency effects that gradually diminished over time. This finding suggests that practicing an incongruent key mapping itself can shape space-time associations, eventually rendering that mapping not-so-incongruent in participants’ minds. Thus, factors that shape the way we think about abstract concepts like time can take place on a variety of time scales. This feature of mental representations contrasts to some extent with the existing literature which often refers to “a representation” of time, perhaps implying that how we conceptualize time is static. However, the dynamic nature of these mental representations is also consistent with other studies that have revealed dynamicity in how we think about time (e.g., [117, 20, 124, 44].

**Broader Impacts**

The studies presented here collectively advance our understanding of the nature of our rich representations of abstract concepts. While the studies focus on mental representations of time, the clarity they provide on the role of linguistic metaphor in
constructing and using those representations can contribute to understanding the role that language plays in our mental representations of structured abstract domains more generally.

The work presented here is basic research (though its lesser-known name of *fundamental* research seems more appropriate to me). As such, it aims to improve scientific theories for understanding and predicting phenomena. However, metaphor is pervasive – for example, it infiltrates political debates, medical communication, and education. This pervasiveness opens many possibilities for downstream application of this knowledge to real-world problems. With a firmer understanding of the role of language – particularly metaphor – in shaping our mental representations of domains like time that we can’t experience first-hand (i.e., through vision or touch), we may eventually be able to harness the power of metaphor to improve the world.
Bibliography


