

**Contributions of Population Stereotypes and Mental Simulations to Sentence
Comprehension**

by

Morgan Teskey
BSc, University of Calgary, 2015

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Morgan Teskey, 2017
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

**Contributions of Population Stereotypes and Mental Simulations to Sentence
Comprehension**

by

Morgan Teskey
BSc, University of Calgary, 2015

Supervisory Committee

Dr. Michael E. J. Masson (Department of Psychology)
Supervisor

Dr. Daniel N. Bub (Department of Psychology)
Co-Supervisor

Abstract

Embodied accounts of action-language processing propose that meaning is constructed with the assistance of relevant sensory-motor representations (eg., Fischer & Zwaan, 2008). In support of this view, comprehending an action-sentence can slow the production of an overt action, when features of that action are incompatible with corresponding sentence features (Glenberg & Kaschak, 2002). Additionally, performing an overt action can impede the comprehension of incompatible action-sentences (Zwaan & Taylor, 2006). Action-sentence comprehension can even be disrupted by watching visual displays with incompatible directional features. Namely, comprehending a sentence describing a movement in a clockwise or counterclockwise direction is less efficient when simultaneously viewing a stimulus moving in an incompatible direction, even when no overt manual rotation action is performed. Embodied accounts contend that such *action-sentence compatibility effects* arise as a result of covert simulations of specific motor programs developed through one's physical experiences with particular objects. I present evidence that these effects could also be generated by a more abstract type of knowledge, that is not tied to a particular object. I am referring here to the idea of a population stereotype, which is the natural tendency of people to associate the direction of certain actions with the conceptual properties of a physical display (e.g., a clockwise device rotation implies an increase in device output). Such population stereotypes typically are consistent with specific motor experiences. For example, turning down the volume of a stereo in many cases involves a counterclockwise rotation of a dial, and this experience is consistent with a population stereotype that implies that reducing a quantity is achieved by a counterclockwise action. If comprehension of a sentence describing reducing the volume on a stereo is faster while turning a dial in a counterclockwise direction, it can not be determined if a resulting compatibility effect reflects compatibility between the described action and the stereotype, or between the described action and real motor experiences. I will present a case in which a population stereotype is not compatible with everyday experiences and establish that population stereotypes make a substantial contribution to action-sentence compatibility effects. I will also report a number of unsuccessful attempts to replicate previous studies of action-sentence compatibility and discuss replication attempts made by others.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
Acknowledgments	vii
1 Introduction	1
1.1 Sentence Direction Judgment Test	9
2 Experiments	
2.1 Experiment 1	13
2.1.1 Method	14
2.1.2 Results	18
2.1.3 Discussion	24
2.2 Experiment 2	26
2.2.1 Method	27
2.2.2 Results	28
2.2.3 Discussion	30
2.3 Experiments 3 and 3a	31
2.3.1 Experiment 3a	31
2.3.1.1 Method	31
2.3.1.2 Results	33
2.3.1.3 Discussion	35
2.3.2 Experiment 3b	35
2.3.2.1 Method	36

2.3.2.2 Results	37
2.3.2.3 Discussion	39
3 Discussion	40
4 References	45
A Critical Stimuli	
A. 1 Sentence Direction Judgment Task	47
A. 2 Experiment 1	48
A. 3 Experiment 2	49
A. 4 Experiment 3a	53
A. 5 Experiment 3b	56

List of Figures

Figure 1.1	Mean percent correct in the Sentence Direction Judgment Task for container and non-container stimuli. Blocks represent successively presented groups of 4 trials. Error bars are 95% confidence intervals	12
Figure 2.1.1	Mean sensibility judgment response times in Experiment 1 for container and non-container stimuli in compatible and incompatible conditions. Error bars are 95% confidence intervals	20
Figure 2.1.2	Mean compatibility effect in Experiment 1 as a function of response-time quintile and sentence type. Each data point is positioned on the horizontal axis according to the mean response time across compatible and incompatible trials in a particular quintile. Error bars represent 95% confidence intervals	22
Figure 2.1.3	Estimates of the compatibility effect for individual sentences relative to the average effect across sentences within each type (container, non-container, and filler) with the average set to zero in each case and corresponding 95% confidence intervals. Sentences are ordered according to their estimated effect sizes	25
Figure 2.2.1	Mean sensibility judgment response times in Experiment 2 for container and non-container stimuli in compatible and incompatible conditions. Error bars are 95% confidence intervals.	31
Figure 2.3.1	Mean reading time per segment in Experiment 3a for container (A) and non-container (B) stimuli, for compatible and incompatible conditions. Error bars are 95% confidence intervals	35
Figure 2.3.2	Mean reading time per segment in Experiment 3b for container (A) and non-container (B) stimuli, for compatible and incompatible conditions. Error bars are 95% confidence intervals	39

Acknowledgements

I would like to thank:

Dr. Michael Masson For his constant mentorship, immense support, and especially for his patient attempts to instill in me his keen attention to detail.

Dr. Daniel Bub For his insightful ideas, enthusiasm for cognition, and especially for keeping my eyes open to the bigger picture in which this project fits.

Marnie Jedynak For her help with programming, data collection, and data wrangling for the experiments reported here.

I would also like to thank the team of research assistants for their assistance with data collection for this project, as well as my fellow grad students in the Cognition and Action Laboratory for their encouragement and comradery.

Chapter 1

Introduction

Embodied accounts of action-language processing propose that language meaning is grounded in sensory-motor representations of past experiences (e.g., Fischer & Zwaan, 2008; Glenberg, 2015). A central tenet of this theory is that action-language evokes activation of neural substrates involved in performing described actions (Bergen, 2015; Jeannerod, 2001). For example, understanding a sentence such as “*The race car driver turns on the ignition*” would depend on activating motor constituents of actions typically made to reach the end-state goal of turning on a car’s engine. Neuroimaging studies have linked the comprehension of action-language to activation of corresponding motor circuitry (e.g., Tettamanti, et al., 2005; Pulvermüller, Härle, & Hummel, 2001; de Vega, et al., 2014). Activation of this nature is commonly referred to as motor resonance or mental simulation, with the latter term favoured in this document. Behavioural evidence in support of a mental simulation view of language processing is largely reliant on demonstrations of action-sentence compatibility effects (see: Meteyard & Vigliocco 2008). Action-sentence compatibility is a measure of interference and/or facilitation between overt action production and action-sentence comprehension (e.g., Borreggine & Kaschak, 2006; Glenberg & Kaschak, 2002). For example, Glenberg and Kaschak (2002) found that sensibility judgments were slowed when the direction of motion required to make a key press response (away or towards the body) was incompatible with the direction of motion implied by a sentence (e.g., “*Andy delivered the pizza to you*” [a *towards* sentence] or “*You delivered the pizza to Andy*” [an *away* sentence]), compared to when the directional features were compatible. This result is consistent with the mental simulation hypothesis, which

would predict that sentence comprehension activates motor representations, which can prime the production of motor responses with compatible features.

Glenberg and Kaschak (2002) propose that sentence comprehension elicits mental simulations in the following way: words and phrases are mapped to perceptual symbols, which are non-arbitrary and based on sensory representations in the brain. When these symbols are activated (by reading or hearing a sentence), they produce affordances, or potential interactions between people and objects (Tucker & Ellis, 1998). For instance, the perceptual symbol of a teacup could produce affordances of grasping the handle, bringing the cup towards the mouth, etc. The content of a sentence directs the activation of perceptual symbols, and the corresponding affordances of sentence components are joined together as a mental simulation. This mental simulation is claimed to facilitate or inhibit the production of overt motor responses, which produces an action-sentence compatibility effect, at least in situations in which subjects can prepare a motor response while comprehending a sentence (see Borreggine & Kaschak, 2006).

In addition to quantifying the effect that comprehending a sentence has on the production of an overt action, studies of action-sentence compatibility effects also examine how performing overt actions can influence sentence comprehension. A study by Zwaan and Taylor (2006) required subjects to turn a dial in a clockwise or counterclockwise direction to progress through segments of a fragmented manual-rotation sentence. These sentences described actions that could be carried out in a clockwise or counterclockwise direction (e.g., “*After disposing of the burnt-out light the projectionist screwed in the new one*” [a clockwise sentence]). Reading speed was slowed when the direction that the dial was turned was incompatible with the direction of action implied by the sentence. However, this effect was present only for the verb-

phrase segment of the sentence (such as “*screwed in*”, in the previous example). The restriction of the action-sentence compatibility effects to the verb-phrase of the sentence provides an indication that the manner in which mental simulations may elicit action-sentence compatibility effects may not be as straightforward as mental simulations priming motor responses. If motor simulations generate action-sentence compatibility effects, then one would expect such effects to be present in all sentence components following the verb, as an online simulation of action would be continually active throughout sentence processing

Zwaan and Taylor put forth two possible hypotheses to account for the localized nature of their result. First, that following the verb-phrase, the reader’s attention shifts from the action to the referent object, and this attentional shift interferes with mental simulation. For example, after reading the verb-phrase in the sentence “*His pencil was dull so before the SAT he sharpens his pencil*”, attention would be drawn away from the clockwise sharpening action, and towards the pencil, which does not have a directional feature because it is the sharpener that is being turned. Although this hypothesis appears plausible in the context of this study, in which sentences were set up such that the referent of the action is introduced before the verb-phrase and is presented again immediately after the verb, it seems less plausible when applied to other sentence structures. For example, if the sentence was modified to read: “*Before the test, he sharpens his pencil*”, then the nature of the mental simulation would now be directly tied to the introduction of the referent (sharpening scissors for instance, would rely on the activation of a different motor program than sharpening a pencil). If mental simulations were disrupted by the introduction of the referent, then sentences like these could not be simulated, as simulation is not possible before the object is introduced. This scenario would directly conflict with the embodied

view of language processing that the authors support, which claims that an inability to simulate the sentence would disrupt sentence comprehension.

The second hypothesis put forth by Zwaan and Taylor to account for the limitation of the action-sentence compatibility effect to the verb-phrase is that such action-sentence compatibility effects are inherently short lived. This hypothesis is at odds with studies such as Glenberg and Kaschak's (2002) sensibility judgment study, in which an action-sentence compatibility effect was produced when a judgment response was made offline, some time following the point at which the directional component of the sentence was clear (see also Zwaan & Taylor, 2006; Experiment 3). To resolve this paradox, Zwaan and Taylor suggest a two-stage model of mental simulations. In the first stage, simulations occur online at the verb-phrase, and then quickly dissipate. In the second stage there is a re-simulation of the action when a judgment about the sentence is made.

An alternative hypothesis to account the localized nature of the action-sentence compatibility effect is that a feature of the verb-phrase, independent of the action simulation, is responsible for producing compatibility effects. If for instance, people have a tendency to associate the outcome of something being turned on with a clockwise direction of movement, then reading the sentence "*The race car driver turns on the ignition*", may produce a compatibility effect at the time of the verb-phrase because it is the only part of the sentence that evoked this association. Relevant motor experiences related to the specific action may be called upon if someone is asked to mentally image the action, or judge a feature of the whole sentence (such as its sensibility). However, compatibility effects produced by passively reading the sentence by turning a dial need not rely on embodied representations of specific actions, but may instead rely on general action-outcome associations. To support this hypothesis, I would need to

find evidence that generic actions (such as a clockwise rotation) can be associated with conceptual outcomes (such as turning on a device). Such evidence is available in the form of *population stereotypes*, and I propose that these population stereotypes are a possible, non-embodied source of the action-sentence compatibility effect.

Population stereotypes reflect the relative frequency at which populations associate certain actions with conceptual outcomes (Proctor and Vu, 2006). Population stereotypes related to an action and its associated consequence have been well established due to their use in engineering and design. For example, understanding that moving a vertical toggle switch in the upright position is typically associated with a device being on would be helpful in designing an intuitive switch mechanism for a particular device (Bergum & Bergum, 1981; Proctor and Vu, 2006). Of particular interest to the present studies is the *clockwise-as-increasing* population stereotype. For instance, clockwise rotation of a dimmer switch connected to a light is associated with an increase in light intensity (Bergum & Bergum, 1981). Likewise, a clockwise rotation of a dial connected to a digital number display is associated with increasing the value of the display (Bergum & Bergum, 1981; Smith, 1981). Clockwise rotation is also associated with turning devices on (Proctor & Vu, 2006). I suggest that a *clockwise-as-increasing* stereotype can account for the compatibility effects among manual-rotation sentences reported by Zwaan and Taylor (2006), which were previously ascribed to simulation-based action-compatibility. As suggested previously, a sentence such as “*The race car driver turned on the ignition*” may show compatibility with a clockwise task due to the tendency to associate the action of a clockwise rotation with an object being “*turned on*”. I now have evidence from a population stereotypes account that this is a reasonable suggestion. The sentence context that would allow for an action to be simulated (including details such as the identity of the object which was turned on) may be

of no consequence to the production of a compatibility effect. Perhaps clockwise compatibility could be seen even for a sentence such as “*The cheetah turned on his coffee maker*”, as the conceptual outcome of a device being turned on is associated with a clockwise action, regardless of the actual means by which the outcome of the action was reached.

The challenge in providing support for a stereotype hypothesis over an embodied view comes from disentangling the relative contributions of population stereotypes and mental simulations, as both factors will generally predict the same behavioural outcomes. For instance, in the previous example of turning on a car’s engine, a simulation view would predict the sentence would show clockwise compatibility due to our experiences of turning a key clockwise in an ignition to turn on an engine; while the stereotype view would predict the same sentence would show clockwise compatibility due to an association between a clockwise turn and the general outcome of something turning *on*. Testing the extent to which population stereotypes contribute to compatibility effects would require a case in which the action’s stereotyped direction competes with the direction of action consistent with real motor experiences. You can demonstrate such a case to yourself by asking which way you twist open a beer bottle. If your first instinct was to answer “clockwise” you have relied on a stereotype and reached an incorrect conclusion. I present evidence that a considerable proportion of people when asked this question will say that they would turn a lid clockwise to open a bottle, despite the wealth of daily experiences people have with opening containers with threaded lids (see Figure 1.1). Zwaan and Taylor (2006) removed the sentence “*Troy twisted open the beer bottle.*” from their sensibility judgment experiments as it did not receive a high counterclockwise rating in a post-experiment questionnaire. The authors presumed this was due to the fact that opening a beer bottle can require a clockwise rotation of the bottle, in addition to the counterclockwise rotation of the lid.

Anecdotally, I have queried many people who erroneously indicated a clockwise direction of motion would be required to open a beer bottle, and these people have indicated they were not, in fact, thinking of how to turn the bottle. I present evidence later in this introduction that this error is made even in a case where rotating the container clockwise to open it is not possible. I suggest instead that this error is a result of the *clockwise-as-increasing* stereotype. Opening a container allows an individual to increase her access to its contents. In this way, the population stereotype of opening a container conflicts with real-world experiences.

This unusual case of opening and closing containers, where population stereotypes act in opposition to real actions, motivated the use of two sentence classes in the present experiments. One class of sentences described an open/close action applied to a container with a threaded lid (*container sentences*). The *non-container* sentence class involved sentences that described manual rotations that were not related to containers, such as unscrewing light bulbs or turning up the volume on a stereo. I used these sentence classes in a series of three experiments to determine whether population stereotypes might contribute to action-sentence compatibility effects.

Initial evidence that population stereotypes related to opening and closing containers may be responsible for producing action-sentence compatibility effects can be found in a study by Claus (2015). This study adapted the methodology introduced by Zwaan and Taylor (2006), which required a dial-turning motion to progress through fragmented sentences. Unlike Zwaan and Taylor, Claus's study utilized only sentences that would fit into my container sentence class. As well, the sentences were presented such that the first word was the name of the actor, followed immediately by the verb-phrase and referent as a single segment (e.g., "*opens the lemonade bottle*"). This provided an opportunity to examine online compatibility effects that

arose while processing information about opening and closing containers. Claus found a significant effect of compatibility on reading speed at the verb-phrase segment of the sentence. Crucially, this effect was in the opposite direction of the compatibility effect found at the verb-phrase in the Zwaan and Taylor (2006) study. That is, reading speeds were faster when the action in the sentence was incompatible with the direction of motion the dial was being turned. This negative compatibility effect can be explained by my population stereotype hypothesis, under which the negative compatibility effect could be reframed as positive compatibility between the overt turning action and the stereotype associated with the outcome of the action. For example, the sentence segment “*opens a lemonade bottle*” was shown by Claus to be comprehended more efficiently when turning a dial in a clockwise direction. As actually opening a lemonade bottle requires the lid to be turned in a counterclockwise direction, there is a processing advantage when the actual action described by the sentence and the overt action required by the task are incompatible. Conversely, the population stereotype for opening a lemonade bottle indicates a clockwise direction of motion matches the outcome of the sentence’s action. Thus, a processing advantage exists when the population stereotype implied by the sentence is compatible with the overt action required by the task. Therefore, I argue that Claus’ finding is consistent with my hypothesis that population stereotypes contribute to the production of compatibility effects.

Based on the results of Zwaan and Taylor (2006) and Claus (2015) I predict three possible outcomes of the present experiments:

1. If mental simulations alone are responsible for producing the action-sentence compatibility effect, then both container and non-container sentences should produce positive compatibility effects. Compatibility is defined as congruency between the direction of movement of a

described action and the directional component of the task (overt or visual motion). Therefore, effects produced by motor representations based on actual movements will show positive compatibility results, such as those reported by Zwaan and Taylor (2006).

2. If population stereotypes alone are responsible for producing action-sentence compatibility effects, then positive compatibility effects would be expected for non-container stimuli, as population stereotypes of this sentence class are associated with the direction of the actual action described by the sentences. However, container stimuli will produce negative compatibility effects similar to those reported by Claus (2015), as the mapping between task and actual action is a reversal of the mapping between the task and the direction associated with the stereotype.

3. If both mental simulations and population stereotypes contribute to action-sentence compatibility effects, then a positive compatibility effect for non-container stimuli would be predicted, and a weak or null effect of compatibility for container stimuli would be expected. For container sentences, the directions of action suggested by the population stereotype and by the mental simulation act in opposition to each other, therefore the effect of each factor may negate the contributions of the opposing factor.

1.1 Sentence Direction Judgment Task

I first needed to establish that the two sentence classes (container and non-container) were in fact construed differently from each other. Upon completion of various experiments, subjects completed a post-test to determine what they believed was the directional component of actions implied by a sentences belonging to the two sentence classes.

Subjects were presented with sentences one-at-a-time on a computer monitor. Subjects read each sentence and were instructed to turn a dial mechanism in front of them in the direction

that matched the direction of the action in the sentence (clockwise or counterclockwise) as quickly and accurately as possible. The dial mechanism was approximately 3.5 cm in diameter and was mounted on an 11 cm x 8 cm x 6 cm box plugged into our testing computer. The dial was located on the top of the box and could be turned clockwise or counter-clockwise relative to the surface of the box. The dial could be turned 90-degrees, with each 10-degree turn sending a letter response to the computer as if a key on a keyboard had been pressed. A set of springs inside the dial mechanism caused it to self-center on release. The tension on the dial was low enough that subjects could turn the dial easily when pinching it between their fingertips, yet high enough that the dial did not continue in the direction of turning due to momentum upon release.

Each subject read 64 critical sentences (32 container and 32 non-container), with half of each sentence class describing clockwise actions and half describing counterclockwise actions. The full set of critical stimuli are listed in Appendix A.1.

227 subjects completed this task. 117 of these subjects completed this task following completion of Experiment 1. Data from those subjects is further described in the results of Experiment 1.

Mean percent correct for all 227 subjects is showing in Figure 1.1. Results of this task show that container items were responded to less accurately (65.1%) than non-container items (87.4%; $F[1, 226] = 123.9, MSE = 455, p < 0.001$). This shows a clear difference in subjects' ability to correctly identify the direction of action described in a sentence, based on my sentence class distinction. As well, I now have evidence that container sentences are not responded to incorrectly due to a simulation of turning the container rather than the lid, as the sentence "*James closes the fuel cap on his car*", in which turning the container is not possible, resulted in an accuracy rating of 71%, which is similar to the other container, but not non-container, sentences.

I do not believe that the Sentence Direction Judgment Task was a direct measure of population stereotypes. Population stereotypes are typically measured by a pencil-and-paper task that queries subjects on their perception of relationships between control movements and depicted devices (Proctor & Vu, 2006). Sentences in this task include context that could potentially impact the subject's responses, compared to when they were responding to a single word or drawing. As well, though each sentence was presented once, the subjects had many exposures to reading about each type of action and had the ability to change their construal of the actions over time. The first presentation of a stimulus of a particular class indeed was responded to less accurately than subsequent stimuli of that class. For container stimuli the first item was responded to 11.7% less accurately than the mean of the following container trials (65.4%; $F [1, 226] = 14.6, MSE = 1062, p < 0.001$). Similarly, the first non-container item was responded to 4.7% less accurately than subsequent non-container items (87.5%), though the statistical evidence for this difference is very weak ($F [1, 226] = 3.9, MSE = 639, p = 0.05$).

This Sentence Direction Judgment Task demonstrates that despite both sentence classes describing common actions that all subjects have likely performed, items in the container sentence class are construed less accurately than those in the non-container class. This error may reflect a stereotype -clockwise for open- that contradicts the actual experience of opening and closing containers. This validates my motivation to examine differences between these sentence classes on measures of action-sentence compatibility to determine if action-sentence compatibility effects are produced, at least in part, due to influences stemming from population

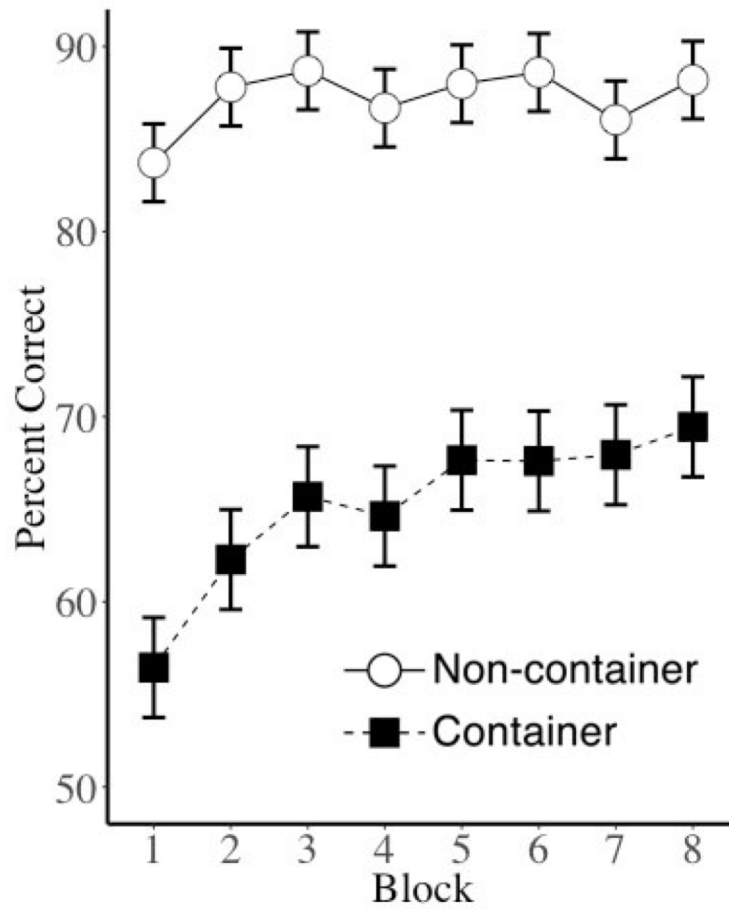


Fig. 1.1. Mean percent correct in the Sentence Direction Judgment Task for container and non-container stimuli. Blocks represent successively presented groups of 4 trials. Error bars are 95% confidence intervals.

stereotypes.

Chapter 2

Experiments

2.1 Experiment 1

The Sentence Judgment Direction Task provided support for my hypothesis that sentence comprehension is influenced, at least in part, by population stereotypes. I believe that inflated error rates for container sentences reflect contributions made by population stereotypes that contradict actual motor experiences of opening and closing containers. On the other hand, non-container sentences are responded to with high accuracy because, in this case, population stereotypes show agreement with real motor experiences. I next asked whether it is the stereotype or motor representation that underlies action-sentence compatibility effects. Both container and non-container sentences were used in an adaptation of methodology from Zwaan and Taylor's (2006) Experiment 3, which measured sensibility judgment time to manual rotation sentences. Although traditional measures of action-sentence compatibility effect examine how action-language can interfere with the production of an overt action (or vice-versa), this paradigm differed in that the actions made in response to stimuli (left and right button presses) were not mapped to a feature of the sentence stimuli (describing manual rotation actions). Rather, compatibility mappings existed between the directional component of a sentence, and the directional component of a rotating visual stimulus. Prior to conducting this study, Zwaan and Taylor (Experiment 1) established that observing a moving visual stimulus could interfere with the production of an incompatible overt action. They also found that sensibility judgments were made more quickly when responses were made by producing a manual rotation response in a direction compatible with the direction implied by the sentence (Experiment 2). Taken together,

these findings served as a methodological basis for their third experiment, which I attempted to replicate in my Experiment 1. Zwaan and Taylor found that sensibility judgments to manual rotation sentences were made 53 ms more quickly when observing a compatible, rather than an incompatible rotating visual stimulus. Their critical stimuli were 16 sentences describing manual rotation actions (8 clockwise, 8 counterclockwise). Two of these stimuli would fit into my container stimulus class (“*Liza opened the pickle jar*” and “*Bob opened the gas tank*”), whereas the other 14 would belong into the non-container stimulus class (eg., “*Erin used the can opener*”). In my version of the study I used 64 critical stimuli (32 for each of sentence classes, 16 of each sentence class describing clockwise actions, 16 describing counterclockwise actions), which allowed me to compare compatibility effects between stimuli where population stereotypes contradict motor experiences (container sentence) and those where population stereotypes show agreement with motor experiences (non-container sentences).

2.1.1 Method

Subjects

One hundred twenty-nine students (101 female; median age = 20 years, ranging from 17 to 33 years) at the University of Victoria participated to earn extra credit in an undergraduate psychology course. The University of Victoria Human Research Ethics Board approved the experiment reported here.

Linguistic Stimuli

Two sets of thirty-two sentences each were constructed. The full set of critical stimuli are listed in Appendix A.2. Container sentences described an action of opening or closing a container with a threaded lid (e.g., water bottle, jam jar, gas cap). Sixteen container sentences described

clockwise (close) actions, and the other 16 described counterclockwise (open) actions. Non-container sentences described manual rotation actions not associated with containers, such as tightening a screw, or turning up the volume on a car radio. Actions that were ambiguous in their implied direction of rotation, such as turning a doorknob or setting a washing machine, which differ across particular cases, we deliberately excluded. Sixteen of the non-container sentences described clockwise actions, and 16 counterclockwise. Some of the sentences used in this test were the same as those used in the Sentence Direction Judgment Task and others were modified versions of those sentences, in which some of the non-essential contextual content was removed or modified.

Seventy sensible filler sentences were also included (6 practice, 64 experimental). Filler sentences did not imply a clockwise or counter-clockwise action (e.g., “Gary worries that he will miss his flight.”, “The red light on the answering machine flashes.”). One hundred thirty-four non-sensible stimuli (6 practice, 128 experimental) were constructed by writing sensible, grammatical sentences, then transforming each sentence through word substitutions until no longer sensible (e.g. “The hostile biographer coughs the telephone awake.”, “The ratios are teasing the bird.”).

All sentences were spoken by a native Canadian English speaker and recorded with the Audacity (version 2.1.1) software. All sentences were presented from a third-person perspective and set in present tense. The spoken duration of the sentences varied in length from 1,600 ms to approximately 3,200 ms.

Visual stimuli

An image of a black cross consisting of two perpendicular lines of equal length (5.2 degrees when viewed from 50 cm) and thickness (0.3 degrees) was constructed. Multiple images of the

cross were created by rotating the original upright image in successive steps of 10-degrees. During the experiment, the sequence of images of the cross were presented for 100 ms each. This display gave the appearance of a cross rotating at a consistent pace of 10-degrees per 100 ms in either the clockwise or counterclockwise direction. On one third of the filler trials (25% of all trials), the cross changed from black to red for 900 ms, then went back to black. The colour change occurred a maximum of one time during a particular trial and was initiated either 500 ms, 1,000 ms, or 1,500 ms after the onset of the trial. On all trials, the rotating cross was continuously in view until the subject made a sensibility judgment with a button press or the auditory sentence was complete, whichever event occurred last.

Design

The 64 critical sentences were randomly divided into two lists of 32 items (16 container and 16 non-container sentences) with an equal number of clockwise and counterclockwise sentences of each type. In addition, a random half of the sensible and non-sensible filler sentences were assigned to each list. One list of sentences was presented in a single block of trials during which the cross rotated in a clockwise direction and the other list was presented in a block with the cross rotating counterclockwise. Assignment of sentence list to cross-turning direction was counterbalanced across subjects, as was the order of presentation of the blocks, which resulted in a total of four counterbalancing conditions.

The planned sample size for the experiment was 60, based on a standard power analysis for significance testing conducted with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). I estimated the required sample size for a target power level of .80, type I error probability of .05, a small effect size ($f = 0.1$), and an assumed correlation between repeated-measures conditions ranging between .80 and .90 (this correlation range was a conservative estimate based on my

lab's prior research with comprehension of action sentences).

In the first version of this experiment ($n = 60$), the assignment of critical sentences to lists was imperfect, resulting in between 6 and 10, rather than 8, sentences of each class (e.g., container-clockwise) assigned to each of the two blocks of trials. Each block still contained the same number of sentences from each sentence type, but numbers of clockwise and counterclockwise sentences was unequal across the two blocks. This uneven counterbalancing was rectified in the second version of the experiment ($n = 69$). The pattern of results in the two versions of the experiment was the same, so I combined the data from both versions and present analyses of the combined data.

Procedure

Subjects were tested individually in a quiet room, supervised by an experimenter. The experiment was controlled by an iMac computer with a separate monitor viewed by the subject. Sentences were presented over headphones and subjects were instructed to classify each one as sensible or non-sensible as quickly and as accurately as possible. They made their responses on a button box (Cedrus Response Pad Series RB-844), which had an upper and lower row of buttons. Subjects rested their index fingers on two buttons in the upper row and their thumbs on two buttons in the lower row. Index fingers were used for sensibility judgments and the assignment of the left or right finger to sensible and non-sensible sentences was counterbalanced across subjects. They were also instructed to watch the rotating black cross on the computer monitor as they listened to the sentences. If they detected a color change in the cross (from black to red), they were to use either thumb to make a button press. The layout of the button box was such that subjects could rest their index fingers and thumbs on the indicated buttons throughout the experiment.

Each block of trials began with 6 practice trials (half sensible fillers and half non-sensible fillers), followed by a randomly ordered sequence of 32 critical and 96 filler sentences. Subjects were given a self-paced rest break following the practice trials, halfway through each block, and after the first block was finished.

Following the experimental procedure subjects completed the Sentence Direction Judgment Task as described in the introduction.

2.1.2 Results

Before analyses were conducted, 12 subjects were excluded from the data set because their accuracy in making either sentence sensibility judgments or detecting the color-change in the rotating cross was below 80%. The data for the remaining 117 subjects were included in the analyses that I report. The mean percent of color changes in the rotating cross that were correctly detected by these 117 subjects was 93.2% ($SD = 4.2$). Their mean accuracy in making sensibility judgments about the critical sentences was 94.5% ($SD = 4.1$).

Data processing and statistical analysis was done using the R statistical language (R Core Team, 2016). Response time (RT) on sensibility judgments was measured from the end of the auditory presentation of the sentence to the button press made by the subject to classify the sentence as sensible or non-sensible. Subjects were free to make a response before the end of a sentence so response times on individual trials were sometimes negative, although trials with values below -250 ms were excluded (0.15% of total trials; see also Zwaan & Taylor, 2006; Exp. 3). Response times longer than 2,500 ms were also excluded as outliers. This upper bound was set so that no more than 0.5% of correct responses were excluded, in keeping with a recommendation by Ulrich and

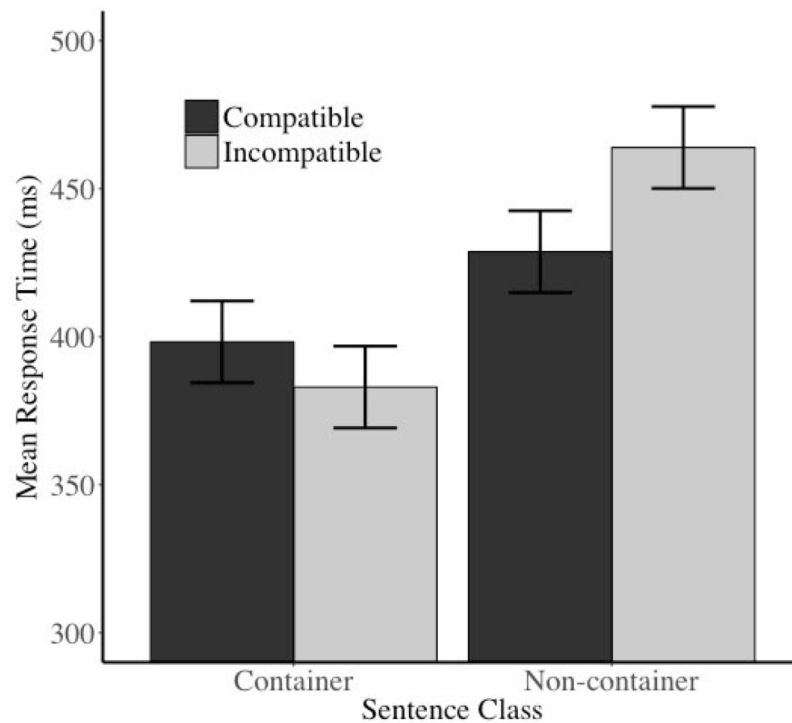


Fig 2.1.1. Mean sensibility judgment response times in Experiment 1 for container and non-container stimuli in compatible and incompatible conditions. Error bars are 95% confidence intervals.

Miller (1994). The mean correct response time for the container and non-container sentences is shown in Figure 2.1.1. An Analysis of Variance (ANOVA) was computed with sentence class and compatibility as a repeated-measures factors which indicated that responses to container sentences were faster than to non-container sentences ($F [1, 116] = 71.54$ $MSE=5,076$, $p<0.001$). This result implies that the content of the container sentences might have been simpler or more familiar. There was no evidence for a main effect of compatibility ($F [1, 116] = 1.75$, $MSE=6,629$, $p=0.19$), but there was evidence of an interaction between sentence class and compatibility ($F [1, 116] = 13.07$, $MSE=5,700$, $p<0.001$). Separate analyses for the two sentence classes indicated the presence of a compatibility effect for non-container sentences ($F [1,116]$

=9.25, $MSE=7,835$, $p<0.01$). There was no compatibility effect found for the container sentences ($F [1,116] =3.03$, $MSE=4,494$, $p=0.08$), though the means show a tendency towards faster responses in the incompatible (383.0 ms) over compatible (398.2 ms) condition.

I computed a similar ANOVA for accuracy on the sensibility judgment task. Sensibility judgments were made more accurately for container stimuli (97.0%) than non-container sentences (92.1%; $F [1, 116] =88.83$, $MSE=32.3$, $p<0.001$). There is no evidence of a speed-accuracy trade-off, as non-container sentences had longer RTs as well as higher errors. Importantly, there was no main effect of compatibility on accuracy ($F [1, 116] = 0.38$, $MSE=31.5$, $p=0.54$), nor was there an interaction between compatibility and sentence class ($F [1, 116] =1.01$, $MSE=30.76$, $p=0.32$).

I also addressed the question of whether the compatibility effect obtained here reflected the contributions of simulations automatically triggered by language. If so, I would expect that even when subjects were particularly fast at accepting a sentence as sensible, the compatibility effect would be evident. Alternatively, if processes contributing to the compatibility effect were engaged only when comprehension was particularly challenging or at a later stage of sentence processing, then the compatibility effect should be most evident on trials with longer response times. This question was examined by computing the compatibility effect at a series of response time quintiles and constructing a delta plot, which shows the effect at each quintile (e.g., Ridderinkhof, 2002). For each subject, the correct response times for a given condition were rank ordered and broken into five equal-sized bins, with the first bin containing the shortest 20% of response times and the final bin containing the longest 20%. The compatibility effect was then computed for each bin for each subject and the mean effect for each bin is shown in Figure 2.1.2. The delta plot for non-container sentences clearly indicates that the effect increases across

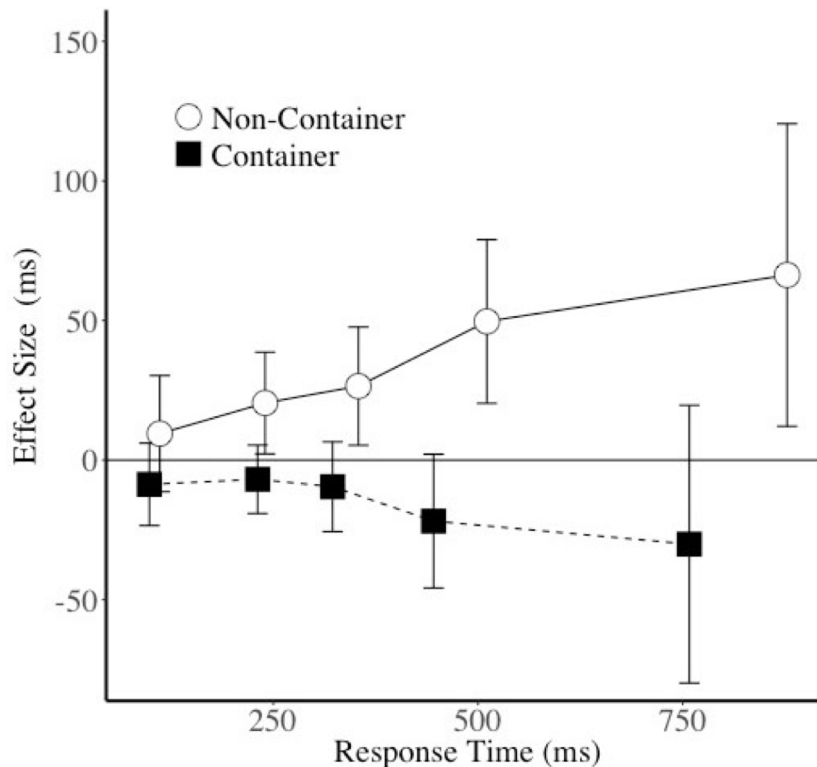


Fig 2.1.2. Mean compatibility effect as a function of response-time quintile and sentence type. Each data point is positioned on the horizontal axis according to the mean response time across compatible and incompatible trials in a particular quintile. Error bars represent 95% confidence intervals.

the response-time distribution. This impression was confirmed by a regression analysis that tested the linear trend for the effect over quintiles ($t=3.321$, $MSE=104.0$, $p>0.001$).

Compatibility effects were not present for container sentences at any quintile.

Data from the Sentence Direction Judgment Task were included in the aggregate data presented in Figure 1.1. For the 117 subjects in Experiment 1, mean accuracy was significantly higher for non-container (88.9%) than for container (69.7%) sentences. This subset of subjects shows the same pattern as the larger group, whose results are presented in the introduction. I then asked whether subjects who are particularly adept at explicitly indicating the correct

direction in which lids of containers must be turned to open or close them might show a different pattern of compatibility effects than was seen with the full sample of 117 subjects. I identified 52 subjects whose accuracy on the Sentence Direction Judgment Task for container sentences was at least 90% (they averaged 94% correct on the non-container sentences). Their sensibility judgment times were analyzed in the same way as the full data set and exactly the same pattern of results was obtained. There was a clear compatibility effect for non-container sentences (52 ms; $F [1, 51] = 13.73, MSE = 5,166, p < 0.001$), but not for container sentences (-12 ms, $F [1, 51] = 0.91, MSE = 4,353, p = 0.35$).

Finally, I addressed the hypothesis that the lack of a compatibility effect with container sentences was the result of two opposing influences, population stereotypes and motor resonance. I reasoned that if two opposing mechanisms were at work, likely varying in their relative degree of strength from trial to trial, then there should be a larger amount of trial-to-trial variability in the sensibility-judgment response times for compatible relative to incompatible trials; on some trials the compatible condition would be favored, on others the incompatible condition. With both mechanisms working toward the same end in the case of non-container sentences, there should be greater consistency in the compatibility effect across trials. This suggestion was tested using a linear mixed-models analysis with the *lmer* function in the *lme4* package in R. A separate analysis was carried out for correct response times to container sentences, to non-container sentences, and to filler sentences. Filler sentences were included because compatibility was an arbitrary designation from sentence to sentence and no meaningful effect could occur. These sentences provided a baseline against which to assess the degree of variability in the other two cases. If the lack of a compatibility effect for container sentences were truly due to the absence of any influence of the direction of rotation in the visual display,

then a similar amount of variability in the compatibility effect should be found for the filler and the container sentences.

For each sentence category, the linear mixed model included a fixed effect of compatibility as well as random effects (intercept and slope) for subjects and for items (sentences). I was particularly interested in the item-by-item estimates of the compatibility effect and the variability of each of those estimates. Greater variation across trials in the compatibility effect for container sentence should yield less stable estimates of the effect for each item, which would be reflected in wider confidence intervals for these estimates.

The results of the linear mixed-models analyses are plotted in Figure 2.1.3 for the three types of sentences. The plot shows the estimated compatibility effect for each sentence, relative to the average effect across sentences (centered at 0) and the 95% confidence interval for each sentence's estimate. These intervals are very similar within each sentence set, but they do vary to a small degree across sentences. It is clear that the compatibility effect is more uniform across sentences and more accurately estimated (smaller confidence intervals) for the non-container sentences than for the container sentences. This was the result expected on the view that while listening to container sentences, observers are subjected to two competing influences with respect to the compatibility between sentence meaning and the rotating cross. The filler sentences show somewhat less variation across sentences in the estimated compatibility effect and much smaller confidence intervals than the container sentences. This contrast suggests that the results seen with container sentences are not likely to be due to a simple absence of an influence of compatibility.

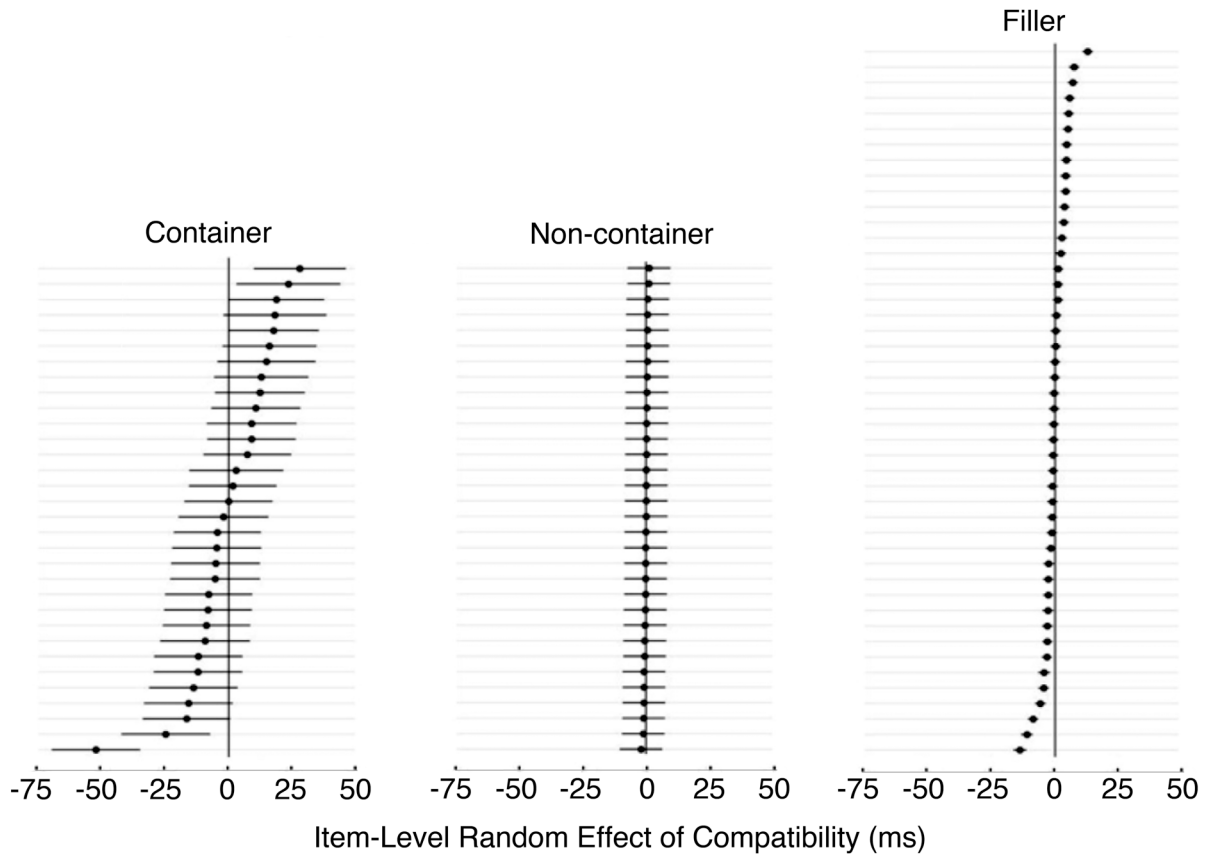


Fig 2.1.3. Estimates of the compatibility effect for individual sentences (y-axis) and corresponding 95% confidence intervals (dark horizontal lines). Estimates are relative to the average effect across sentences within each type (container, non-container, and filler) with the average set to zero in each case. There are 32 sentences each in the container and non-container types, and 46 filler sentences (only sentences presented on trials where the cross did not change color are included). Sentences are ordered according to their estimated effect sizes.

2.1.3 Discussion

I hypothesized that if both population stereotypes and mental simulations contribute to sentence comprehension, then in the case where the two factors are in agreement (the non-container condition) a positive action-sentence compatibility effect would be measured, and in the case where population stereotypes and simulations act in opposition to each other (the container

condition) an action-sentence compatibility effect would not be present. Experiment 1 produced a reliable compatibility effect of approximately 35 ms for the non-container condition, whereas the container condition's -12 ms effect was not statistically reliable, suggesting that action-sentence compatibility effects arise as a result of both population stereotypes and mental simulations. I anticipated that a null compatibility effect in the container condition would appear as a result of influence from two competing sources (mental simulations and population stereotypes). Results of a linear mixed-models analysis of item-level compatibility effect variability provides support for this view.

Claus (2015) found evidence of an online negative action-sentence compatibility effect, which I suggest arose primarily as a result of population stereotypes related to open/close actions. Although container sentences in Experiment 1 showed a trend towards negative compatibility, the effect was not found to be significant, despite having enough power to detect an effect size of a similar magnitude to the effect present in the non-container condition (based on a post hoc analysis using G*Power, power=0.98 assuming, type I error = .05, and effect size $[dz] = 0.37$).

A possible explanation as to why the container sentences did not show a negative compatibility effect similar to that reported by Claus, is the offline nature of the sensibility judgment measure. I hypothesize that by the time a sensibility judgment is made, sentence components peripheral to the action are able to correct the negative influence of the population stereotype. For example upon hearing the sentence "*Jeff opens the apple juice bottle*", the influence of the population stereotype (in a clockwise direction) might be strongest when listening to the verb, but by the time the sentence is judged as sensible, representations related to motor experiences of opening a juice bottle (in a counterclockwise direction) may be called upon

and counteract the compatibility effect produced by the population stereotype. This hypothesis is supported by the results of the linear mixed-models analysis, which shows evidence of conflict between two opposing forces when making a sensibility judgment to container sentences. This hypothesis is further supported by the RT distribution analysis for the non-container sentences. In this case, the influence of the population stereotype would not be counteracted by motor representations. Instead, the influence of the stereotype would persist throughout sentence processing, and motor representations from later sentence components may supplement the compatibility effect that is present at the time of the sensibility judgment, leading to an increasing effect size with longer response times. In contrast, Claus measured the compatibility effect produced by the verb/noun phrase online, and at the point of the verb/noun phrase the influence of the population stereotype might not have yet been counteracted by motor representations associated with real motor experiences.

2.2 Experiment 2

A sensibility-judgment task, such as that in Experiment 1, requires subjects to listen to all or most of a sentence before they are able to make a correct judgment and response. It is therefore not possible to discern from these data at which point in the sentence compatibility effects were produced. My hypothesis is that population stereotypes associated with the directional component of the sentence, as indicated by the verb-phrase, make a large contribution to action-sentence compatibility effects. To gain a clearer understanding of the role of the verb-phrase in producing such compatibility effects I adapted the procedure from Experiment 1 to a passive listening task, which required subjects to respond to an unrelated stimulus at a time directly following a verb-phrase utterance.

2.2.1 Method

Subjects

Thirty students (25 female; median age = 20 years, ranging from 18 to 32 years), at the University of Victoria participated to earn extra credit in an undergraduate psychology course.

Linguistic Stimuli

Sentences were expanded versions of the critical and sensible filler stimuli used in Experiment 1, which were longer in length and contained more contextual information. Critical sentences introduced the referent before the verb phrase, so by the time of verb-phrase offset, the action implied by the sentence was evident (e.g., “His bottle of root beer is beginning to go flat, so he closes the bottle to trap in the bubbles”). Critical stimuli are listed in Appendix A.3. Spoken sentences varied in length from 4,400 ms to approximately 7,800 ms.

Probe questions were presented visually after 25% of trials, and consisted of true or false questions about that trial’s auditory sentence (eg., True or false: *The truck turned on to Pear Street.*).

Design

Critical stimuli were randomly divided into two lists of 32 items (16 container and 16 non-container) with half of each sentence class containing an equal number of clockwise and counterclockwise sentences. Each of the two blocks of the experiment contained stimuli from one of the two lists. Block presentation order was counterbalanced across subjects. Each trial was assigned a colour designation (blue or green), which corresponded to the colour of a visual cue presented during that trial. Within each block half the trials were blue, and the other half green. As well, an equal number of critical sentence/direction types were assigned to each colour. Sentence-colour assignment was counterbalanced across subjects. Each block started

with eight practice trials, followed by 64 experimental trials (32 critical and 32 filler). A self-paced rest break was given between blocks, as well as within each block after following the practice period and halfway through the critical trials.

Procedure

Subjects listened to sentences presented over headphones, and were instructed to listen to the sentence in anticipation of responding to a true/false question about the sentence. At the same time, they focused on a fixation cross in the middle of the screen to watching for the appearance of a coloured square. Upon the appearance of the square, subjects were instructed to make a speeded dial turn in a direction (clockwise or counterclockwise) assigned to the colour (blue or green). They made their responses using the same dial mechanism that was utilized in our Sentence Direction Judgment Task. Colour-direction mappings were maintained for each subject, and counterbalanced across subjects. On critical trials, colour blocks appeared at the end of the verb phrase utterance. On filler trials the colour block SOA was selected randomly from a uniform distribution within the range of critical-stimuli SOAs, with the restriction that coloured square always appeared before the end of the audio sentence. Colour blocks remained on screen until a turning response was made.

Following 25% of trials a true/false probe question was presented on the screen. Subjects made verbal responses to probe questions and responses were scored by an experimenter.

2.2.2 Results

Before analyses were conducted, two subjects were excluded because their accuracy in response to the probe questions was less than 80%. The data from the remaining 28 subjects were included in the analyses I report.

Response time (RT) was recorded from the onset of the colour block until the dial was turned 10 degrees to trigger a response. Response times of less than 100 ms (0.05% trials) or greater than 2,700 ms were removed from analysis, with the upper boundary set such that no more than 0.5% of observations were excluded. Compatibility is defined as congruency between the colour block's associated rotation direction, and the direction of the actual manual rotation direction required to physically carry out the action described in the sentence.

An ANOVA of these data, with sentence class and compatibility as repeated-measures factors indicated that there was no significant effect of either sentence class ($F [1, 27] < 1$, $MSE=10,662$, $p=0.62$), or compatibility ($F [1, 27] = 1.00$, $MSE=9,612$, $p=0.33$). Nor did the two factors interact ($F [1, 27] < 1$, $MSE=2,416$, $p=0.74$). The means of each sentence class and compatibility condition are presented in Figure 2.2.1.

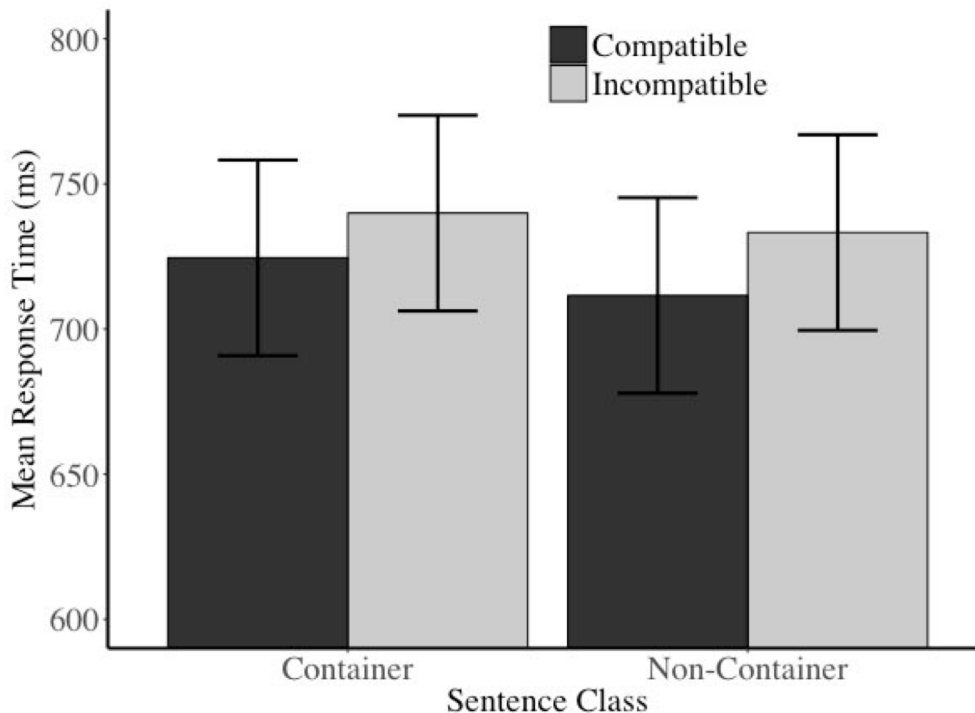


Fig. 2.2.1. Mean sensibility judgment response times in Experiment 2 for *container* and *non-container* stimuli in compatible and incompatible conditions. Error bars are 95% confidence intervals

A similar analysis was applied to the accuracy of the colour-square responses.

Significantly more errors were made on non-container trials (10.1%) than container trials (5.7%) ($F [1,27] = 56.8, MSE=9.8, p<0.001$) which may have indicated the container sentences were simpler to comprehend, and thus required less attention to be paid to the sentence. Importantly there was no significant effect of compatibility on error rates ($F [1,27] <1, MSE=62, p=0.87$), nor did sentence class interact with compatibility ($F [1,27] <1, MSE= 43, p= 0.48$).

2.2.3 Discussion

I was unable to find evidence of an action-sentence compatibility effect for either sentence class, even with the response cue occurring directly following the verb phrase. A post hoc power

analysis suggests this experiment did not have sufficient power to detect an effect (power level of 0.41, type I error probability of .05, effect size (d_z) = 0.34). Further discussion of null action-sentence compatibility effects is presented in the general discussion.

2.3 Experiments 3a and 3b

Experiment's 3a and 3b are replications of the dial-turning-to-read paradigm used in Zwaan and Taylor's (2006) Experiment 4, and by Claus (2015). Experiment 3a was my original attempt to replicate the experiment, while 3b involved slight modifications in an attempt to increase the sensitivity of the measure.

2.3.1 Experiment 3a

2.3.1.1 Method

Subjects

Thirty-five students (27 female; median age = 20 years, ranging from 18 to 41 years), at the University of Victoria participated to earn extra credit in an undergraduate psychology course.

Linguistic Stimuli

The stimuli used in Experiment 2 were modified for Experiment 3a in such a way that each sentence could be divided into seven segments and the verb phrase would occupy the fifth segment. Sentences ranged in length from 12 to 29 words. Segments ranged in length from 1-6 words. For example: "*To assemble/the bookcase/Kailey uses/ a little wrench to/tighten/the screw into/the wooden frame*", where slashes represent the segment boundaries. Critical stimuli are listed in Appendix C.

Probe sentences consisted of open-ended questions about the sentence (eg., “What was Kailey building?”).

Design

The 64 critical sentences were randomly divided into two lists of 32 items (16 container and 16 non-container) with an equal number of clockwise and counterclockwise sentences of each type. In addition, a random half of the filler sentences were assigned to each list. One list of sentences was presented in a single block of trials during which a dial mechanism (the same device used in the Sentence Direction Judgment Task) needed to be rotated in a clockwise direction to progress through the sentence, and the other list was presented in a block requiring counterclockwise rotation. Assignment of sentence list to dial-turning direction was counterbalanced across subjects, as was the order of presentation of the blocks, which resulted in a total of four counterbalancing conditions.

Procedure

Subjects progressed through a fragmented sentence by continuously turning a dial. Each trial began with a fixation cross that remained on the screen until subjects moved the dial in the correct direction. Each 10-degrees of movement triggered the presentation of the next sentence fragment in place of the previous segment (70-degrees of total rotation per sentence). Subjects were instructed to turn the dial at a pace that allowed them to read at their normal reading speed. Upon reaching the end of the sentence (a blank screen or a probe question screen), subjects released the dial, which, upon return to center, triggered the fixation cross for the next trial.

Turning direction was determined by the block and switched after the first block. Eight practice trials were presented at the start of each block followed by 64 experimental trials (32 critical, 32 filler).

Probe questions were presented following 25% of trials and subjects made verbal responses that were scored by the experimenter.

2.3.1.2 Results

Before analyses were conducted, 4 subjects were excluded from the data set because their accuracy in responding to the probe questions was below 80%. The data from the remaining 31 subjects were included in the reported analyses.

Reading time was recorded for each sentence segment, starting from the first appearance of the segment content on the screen, until the dial was turned 10-degrees, at which point the next sentence segment appeared. Reading times of less than 100 ms (0.35% of trials) or greater than 3,000 ms were removed from analysis. This upper RT boundary was established so that no more than 0.5% of observations were excluded.

The mean correct response time for the container and non-container sentences is shown in Figure 2.3.1. An ANOVA was computed with sentence class, sentence segment, and compatibility as a repeated-measures factors which indicated that reading speeds were faster for non-container than for container stimuli ($F [1, 30] = 4.75, MSE = 15,582, p = 0.37$). This difference could be due to differences in sentence length or difficulty. There was a main effect of sentence segment ($F [6, 180] = 30.14, MSE = 41,027, p < 0.001$), which likely reflects differences in segment length. There was no significant effect of compatibility ($F [1, 30] < 1, MSE = 16,664, p = 0.52$), nor did compatibility interact with sentence class ($F [1, 30] < 1, MSE = 10,175, p = 0.58$) or sentence segment ($F [6, 180] < 1, MSE = 62,53, p = 0.07$).

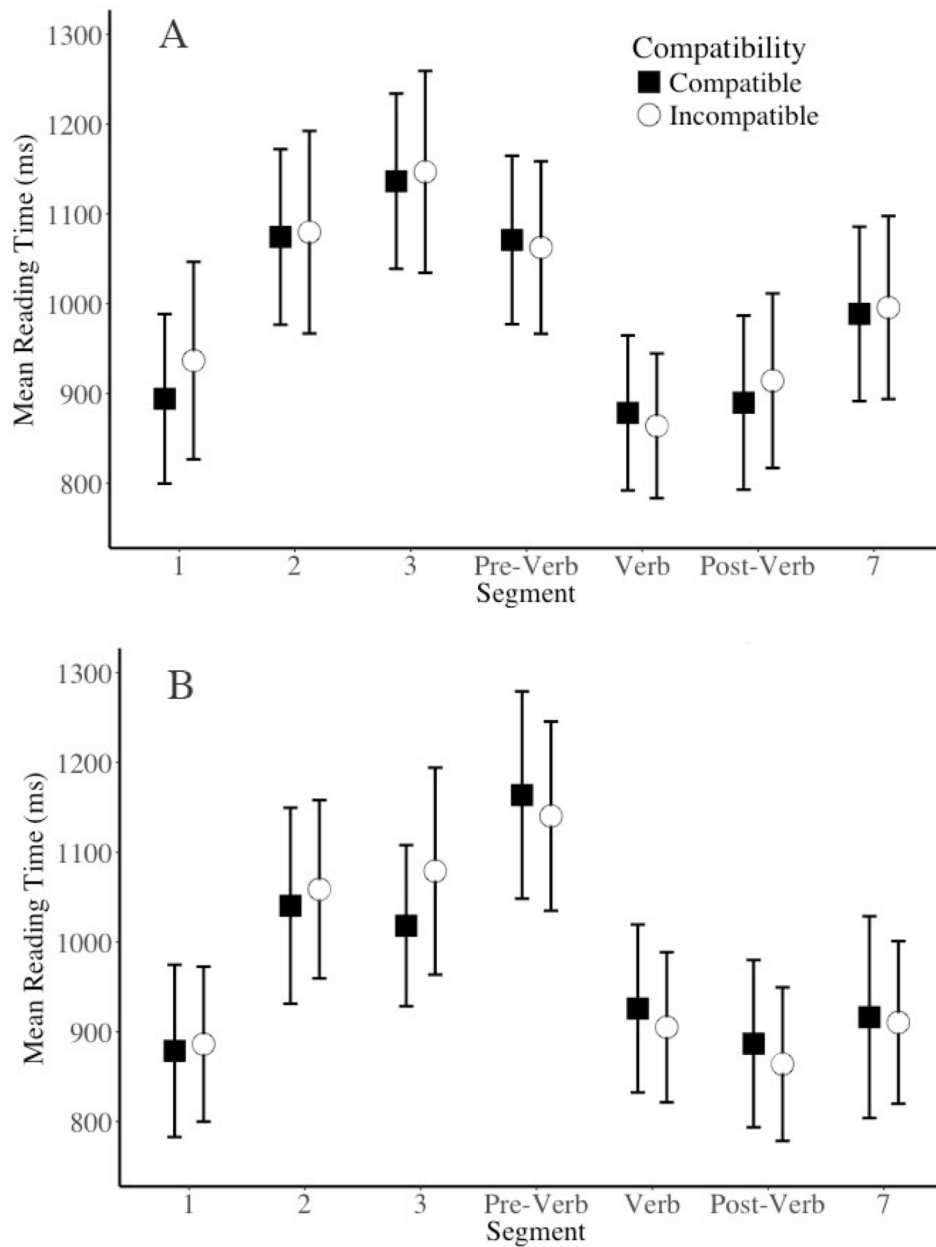


Fig. 2.3.1. Mean reading time per segment in Experiment 3a for container (A) and non-container (B) stimuli, for compatible and incompatible conditions. Error bars are 95% confidence intervals

Next, a separate ANOVA was conducted, which was restricted to the verb-phrase segment of the stimuli, as this is where compatibility effects were present in Zwaan and Taylor's study that utilized the same methodology, as well as by Claus (2015). The verb phrase segment

showed no reliable effect of compatibility ($F [1, 30] = 1.25, MSE=7,714, p=0.27$), nor did compatibility interact with sentence class ($F [1, 30] < 1, MSE=5,396, p=0.81$).

I further restricted the analysis to the verb segment of non-container sentences to match the conditions of the Zwaan and Taylor experiment as closely as possible, as the majority of their stimuli would fall into my non-container sentence class. No evidence for an action-sentence compatibility effect was found ($F [1, 30] < 1, MSE=7,353, p=0.35$).

2.3.1.3 Discussion

I did not replicate the findings of Zwaan and Taylor's (2006) reading speed Experiment. No effect of compatibility on reading speed was found for either sentence class, even when the analysis was restricted to the verb-phrase segment of the sentence. A notable difference between my results and Zwaan and Taylor's is that the average segment reading time in my study was about three times longer than Zwaan and Taylor's. I suspect this is due to the relatively larger amount of words in each of my segments. These longer reading times may have impeded my ability to detect interference on sentence processing time caused by making an incompatible turning action.

I did not have sufficient power to detect an effect of the same magnitude as the effect found by Zwaan and Taylor in the verb-phrase sentence segment (power=0.19). I increased the sample size in version 3b of this study, in an effort to increase power and detect an effect of a similar size to that reported by Zwaan and Taylor. As well, sentence segment lengths were modified to be the same length as those used by Zwaan and Taylor.

2.3.2 Experiment 3b

2.3.2.1 Method

Subjects

Sixty-two students (27 female; median age = 20 years, ranging from 18 to 41 years), at the University of Victoria participated to earn extra credit in an undergraduate psychology course.

Linguistic stimuli

Sentence stimuli were modified from Experiment 3a, such that each sentence contained 11 segments. Sentences ranged in length from 15 to 22 words and all segments were 1-2 words in length. For example: “*The smoke/detector/starts beeping/in the/middle of/the night/and Jim/tiredly/unscrews/it from/the ceiling*”. The restriction of segments to 1-2 words in length was made in an effort to decrease the reading time for each segment. As a result of this, the verb phrase was no longer held in a constant position, and now appeared as early as the fifth segment or as late as the tenth segment.

Probe questions were changed from open ended questions to true/false questions about the sentence (eg., “True or False: Jim was unscrewing a shower head?”).

Apparatus

A dial mechanism with similar specifications as the one used in Experiments 1, 2, and 3a was used, however this dial was not spring loaded. The dial did not return to a set point upon release, but moved continuously in both directions. Resistance on the dial was lower and turning movements required less effort. This change was necessary as the previous dial allowed only a 90-degree rotation in either direction, which would not work with the present Experiment’s 11-segment stimuli.

Design and Procedure

The design and procedure are the same as Experiment 3a, with the exception that an experimenter key press prompted the start of each trial, as the release of the dial could no longer trigger an event.

2.3.2.2. Results

Before analyses were conducted, two subjects were excluded from the data set because their accuracy in responding to the probe questions was below 80%. The data from the remaining 60 subjects were included in the analyses we report. Reading times of less than 100 ms (0.18% of trials) or greater than 1,400 ms were removed from the analysis.

The mean correct response time for the container and non-container sentences is shown in Figure 2.3.2. An ANOVA was computed with sentence class, sentence segment, and compatibility as repeated-measures factors. As the position of the verb phrase was not consistent across trials, segments were either tagged as *pre-verb* (occurring directly before the verb phrase), *verb*, *post-verb* (occurring directly after the presentation of the verb-phrase), or *other* (any other position not directly surrounding the verb segment). Main effects of sentence type and segment were present, but these effects are not informative ($F [1,59] =40.0, MSE=1,652, p<0.001$ and $F [3, 117] =24.89, MSE=1,360, p<0.001$, respectively). There was no significant main effect of compatibility ($F [1,59] <1, MSE=1,325, p=0.62$), nor did compatibility interact with sentence class ($F [1,59] <1, MSE=1,247, p=0.83$) or sentence segment ($F [3, 177] =1.6, MSE=906, p=0.19$).

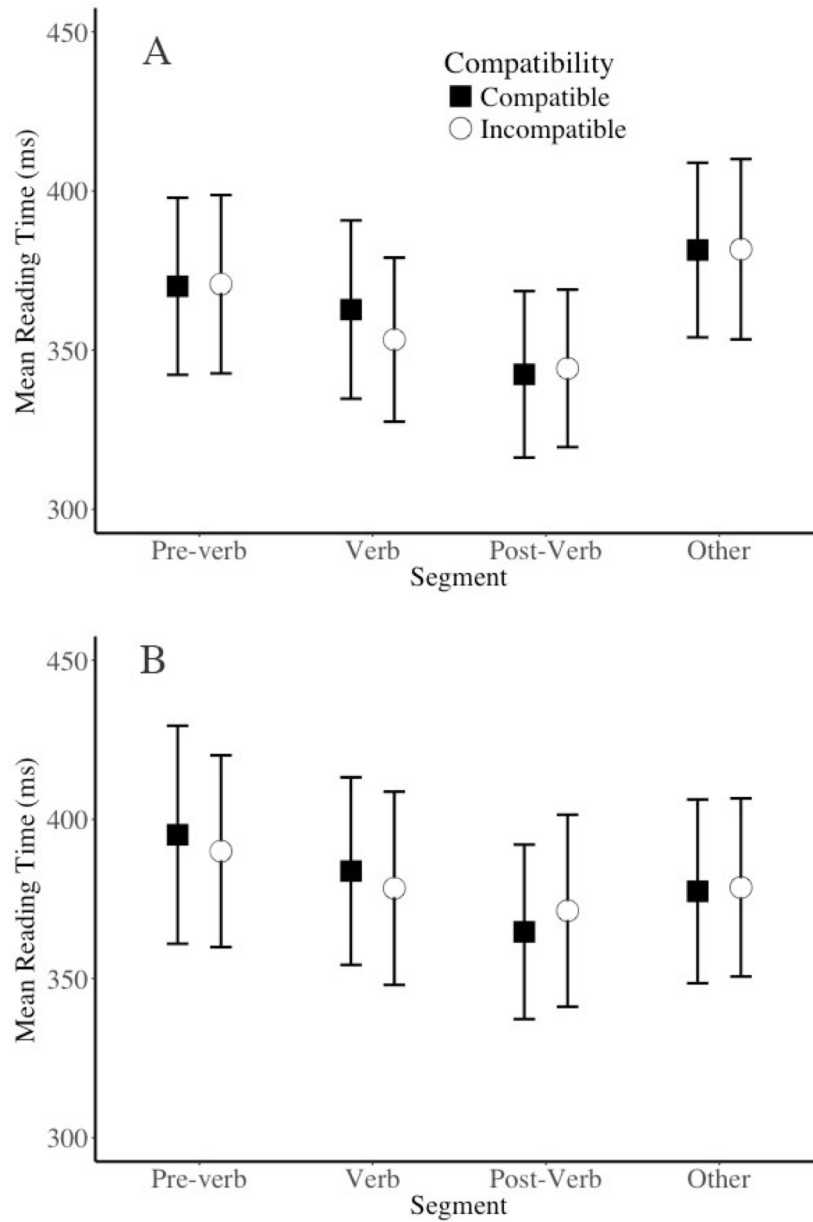


Fig. 2.3.2. Mean reading time per segment in Experiment 3b for container (A) and non-container (B) stimuli, for compatible and incompatible conditions. Error bars are 95% confidence intervals.

Next I restricted the analyses to the verb segment of the sentence. The verb phrase segment showed no reliable effect of compatibility ($F [1, 59] = 3.26, MSE=1019, p=0.08$), nor did compatibility interact with sentence class ($F [1, 30] < 1, MSE=1103, p=0.62$).

Even when analysis of the verb segment was limited to the non-container sentence class, no evidence for an action-sentence compatibility effect was found ($F [1, 59] < 1$, $MSE=1256$, $p=0.41$)

2.3.2.3 Discussion

The changes to the stimuli were successful in making my mean reading times similar to the reading speeds observed by Zwaan and Taylor (2006). This study now had sufficient power to detect an effect of the magnitude reported by Zwaan and Taylor (power=0.95). However, I was still unable to find evidence of an action-sentence compatibility effect, even when the analysis was restricted to the verb phrase of non-container sentences.

Chapter 3

General Discussion

Defining the relative contributions that mental simulations and population stereotypes make to sentence comprehension is important if we are to better understand the extent to which language meaning is grounded in the body and in the physical environment. Action-sentence compatibility effects are thought by some researchers to reflect priming effects between motor activation and action components of a sentence (e.g., Glenberg & Kaschak 2002; Zwaan & Taylor, 2006). I present an alternative hypothesis, positing that action-sentence compatibility effects may arise as a consequence of population stereotypes, which reflect associations between motor features and conceptual action-outcomes (see Procter & Vu, 2006). For example, a clockwise rotation of a dial is generally associated with an increase of some output such as volume or light (Bergum & Bergum, 1981). I suggest that comprehending sentences that describe an increase in some output will therefore be facilitated by a clockwise movement due to population stereotypes, and not due to the activation of specific motor representations corresponding to the described action.

I have evidence in support of my hypothesis from Claus (2015), who found a negative action-sentence compatibility effect at the time that sentence segments referring to opening or closing containers were processed. Though the author presented this evidence in support of an embodied interpretation, I believe that the stimuli used in this experiment present a special case in which directional features stereotypically associated with opening and closing containers act in opposition to motor experiences. For example, if opening a jar is analogous to increasing access to its contents, then opening may be associated with a clockwise movement, rather than

with a counterclockwise movement that would actually be used to open the jar. Under this hypothesis, the negative compatibility effect reported by Claus can be reframed as a positive compatibility effect between the motor action performed by the subject and the population stereotypes related to the verbs in her open/close sentences.

I used a task in which subjects were required to explicitly identify the directional components of action sentences, to establish that it is more difficult to judge the directional component of sentences describing opening/closing containers, compared to other clockwise and counterclockwise actions. Subjects judged the direction of motion implied by critical sentences, and responded to container sentences with only 65.1% accuracy, 22.3% less than non-container sentences (and only 15% above chance). These data give an indication that judging the direction of motion implied by a container sentence is subject to influences other than the recollection of daily experiences. Errors in this task show that sentence construal is likely influenced by population stereotypes, which, in the case of the container sentences, conflict with actual motor experiences. This finding motivated the use of two sentence classes (container and non-container) throughout my experiments.

Across three experiments I adapted methods that were previously successful in measuring action-sentence compatibility effects. All of my experiments involved the presentation of manual-rotation sentences in conjunction with a task element that included a rotational component (visual stimulus or overt motor action). In Experiment 1, I was successful in detecting an action-sentence compatibility effect. I found that in the case of non-container sentences, sensibility judgments were made more quickly when viewing a compatible, rather than incompatible rotating cross. However, the container stimuli did not produce a compatibility effect, though mean response time did show a tendency for judgments in the incompatible

condition to be made more quickly than in the compatible condition. Subjects in Experiment 1 who were minimally influenced by population stereotypes in the Sentence Judgment Direction Task demonstrated this same overall pattern of results in Experiment 1 as those with lower container sentence accuracy. This shows that even though people have the ability to make correct overt judgments about actions that overrule the population stereotypes, those population stereotypes still have a covert influence on subject's abilities to judge sentence meaning.

I infer that population stereotypes and action representations jointly contribute to sentence comprehension, and to action-sentence compatibility effects. The response time distribution analysis in Experiment 1 shows that the combined impact of these two factors accrues slowly, as evidenced by the linearly increasing compatibility effect across the reaction-time distribution for non-container sentences. Slower responses allow more time for action representations to influence performance, an outcome that is not consistent with the claim that motor representations play an obligatory role in sentence comprehension (see Bub, Masson, & Kumar, 2017, for an example of motor compatibility effects that remain invariant across the reaction-time distribution). Another possibility is that when the sentence is more difficult to comprehend (and judge) people will call upon mental simulations as a resource to aid in understanding.

I suggest that the null effect found for container sentences in Experiment 1 emerges as a result of conflict between a population stereotype (driving compatibility effects in one direction), and previous motor experiences, (driving compatibility effects in the opposite direction). Evidence for this conflict comes from the results of a linear mixed-models analysis, which showed a high degree of uncertainty in the estimation of compatibility effects for container sentences compared to non-container and filler sentences.

No compatibility effects were detected in Experiments 2, 3a or 3b. Non-replications of the action-sentence compatibility effect are not a phenomenon unique to my lab. Papesh (2015) attempted to replicate past findings of action-sentence compatibility effects across eight experiments and was unable to find convincing evidence of such an effect across various task conditions and using various response modalities and stimuli, including those used in the original Glenberg and Kaschak (2002) study. In cases where Papesh did find evidence of the action-sentence compatibility effect, it was only for sentences that described an action moving away from the observer. For sentences that described actions moving away from the observer the effect was reversed.

Additionally, Papesh computed *Scales JZS Bayes factor* values for prior studies reporting action-sentence compatibility effects to determine the strength of evidence reported in favour of action-sentence compatibility facilitation. Of the 20 experiments she examined, more than half showed greater support for the null hypothesis, over the alternative hypothesis that an action-sentence compatibility effect was present. Most relevant to my experiments, is the finding that Zwaan and Taylor's (2006) verb region compatibility effect in the turning-to-read study (that I attempted to replicate in Experiments 3a and 3b) showed weak evidence that the null was true over the alternative hypothesis. This suggests that my inability to replicate is not surprising, as evidence for the effect is not strong even in Zwaan and Taylor's data. The only experiment from Zwaan and Taylor (2006) for which there was strong evidence of an effect was their rotating cross/sensibility judgment task (6:1 odds, in favour of finding an effect over the null), which I successfully replicated in the non-container condition in Experiment 1. I also computed a Bayes factor analysis for the observed compatibility effect and found evidence in favor of the

effect (8:1). Though, as discussed above, this evidence in favour of a compatibility effect is not necessarily evidence of mental simulations as Zwaan and Taylor would suggest.

The evidence from Experiment 1, as well as the null results from Experiments 2, 3a, and 3b hold implications for not only the interpretation of action-sentence compatibility effects, but also for the embodied view of language processing as a whole. If language meaning were derived from the activation of relevant sensory-motor representations, then one would expect compatible motor responses to be facilitated at all points of a sentence where simulation was possible. Instead we see that action-sentence compatibility effects are restricted to the verb-phrase (Zwaan & Taylor, 2006; Claus, 2015), and can even present in the opposite direction that would be predicted by a simulation view (Claus, 2015). My results provide evidence that population stereotypes make a substantial contribution to action-sentence compatibility effects. This finding offers an alternative account of empirical findings taken as evidence for embodied theories of language processing, and can reconcile the contradictory nature of past research (Zwaan & Taylor, 2006 versus Claus, 2015). My account suggests that action-language comprehension may not rely entirely on mental simulations, but in addition is influenced by more abstract conceptual knowledge, such as population stereotypes.

References

- Bergen, B. (2015). Embodiment, simulation and meaning. *The Routledge Handbook of Semantics*, 142.
- Bergum, B. O., & Bergum, J. E. (1981). Population stereotypes: An attempt to measure and define. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 25, No. 1, pp. 662-665). Sage Publications.
- Borreggine, K. L., & Kaschak, M. P. (2006). The action–sentence compatibility effect: It's all in the timing. *Cognitive Science*, 30(6), 1097-1112.
- Bub, D. N., Masson, M. E., & Kumar, R. (2017). Time Course of Motor Affordances Evoked by Pictured Objects and Words. *Journal of experimental psychology. Human perception and performance*.
- Claus, B. (2015). Verb gapping: An action-gap compatibility study. *Acta psychologica*, 156, 104-113.
- de Vega, M., León, I., Hernández, J. A., Valdés, M., Padrón, I., & Ferstl, E. C. (2014). Action sentences activate sensory motor regions in the brain independently of their status of reality. *Journal of cognitive neuroscience*, 26(7), 1363-1376.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61(6), 825-850.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin & review*, 9(3), 558-565.
- Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 69(2), 172.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, 14(1), S103-S109.
- Meteyard, L., & Vigliocco, G. (2008). The role of sensory and motor information in semantic representation: A review. *Handbook of cognitive science: An embodied approach*, 293-312.
- Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, 144(6), e116.
- Proctor, R. W., & Vu, K. P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. CRC Press.

- Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and language*, 78(2), 143-168.
- R Core Team (2016). R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Smith, S. L. (1981). Exploring compatibility with words and pictures. *Human Factors*, 23(3), 305-315.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., ... & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of cognitive neuroscience*, 17(2), 273-281.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, 24(3), 830.
- Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, 123(1), 34.
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135(1), 1.

Appendix A

Critical Stimuli

A.1 Critical Stimuli used in the Sentence Direction Judgment Task

Container (Clockwise)

Dave closes the tub of peanut butter.
 Hailey closes the top of the cola bottle.
 He closes the lid on his root beer.
 He closes the mason jar quickly.
 He closes the relish jar.
 James closes the fuel cap on his car.
 Jeff closes the bottle's lid.
 Megan closes the cap on the medicine tube.
 Rick closes the lid on the oil bottle.
 She closes her water bottle.
 She closes the lid on the nail polish.
 She closes the lid on the old jar.
 She closes the top of her thermos quickly.
 She closes the top of the cola bottle.
 The chef closes the glass jar.
 The mother closes the lid on the juice bottle.

Non-container (Clockwise)

Before the race the driver turns on his car.
 For daylight savings he moves the clocks hour hand forward.
 Ginger screws a new light bulb into the lamp.
 He exited the highway.
 He sharpens the pencil.
 He turns the combination lock dial to right-17.
 He turns the dimmer switch to make the light

Container (Counterclockwise)

Greg opens the peanut butter jar.
 Gwen opens the top of the fuel can.
 He opens the pickle jar for his wife.
 He opens the tub of lotion.
 He twists open the cap of the vitamin bottle.
 Jenny opens the lid to the thermos.
 Lex opens the jar of goodies.
 Lucy opens the lid of the jam jar.
 She forcefully opens the opaque jar.
 She opens the bottle of lemonade.
 She opens the jar of almond butter.
 She opens the jar of hazelnut spread.
 She opens the jar of preserves.
 She twists open the cap on the toothpaste tube.
 She twists open the soap bottle.
 Zach opens the plastic water bottle.

Non-container (Counterclockwise)

Adrian removes the screws from the faceplate.
 Beth turns down the radio volume.
 Dan turns back the hour hand on his clock.
 He turned down his stereo.
 He turns off his car engine.
 He turns the lights on the dimmer switch down.
 He unscrews the lugnut from her tire.

brighter.

He turns the key in the ignition to turn on the engine.

Jerry screws in a cable to the television.

Kailey tightens the screws with a small wrench.

Martin tightens the bolt with a wrench.

Sebastian tightens the loose screw on his glasses.

She tightens the screws in the board.

She turns up the volume on the car radio.

She winds forward the egg timer to three minutes.

She winds forward the time on her slow watch.

He unscrews the screw on his laptop.

Jared unscrews the old showerhead.

Jim unscrews the beeping smoke detector.

Kevin turns off the oven.

Luis unscrews the burnt-out light bulb.

She turns back the time on her watch.

She turns down the volume knob on the radio.

The driver turns left on Bay Street.

The professor unscrewed a bulb from the projector.

A.2 Critical Stimuli used in Experiment 1

Container (Clockwise)

He closes the new lid onto the jar

Harvey closes the relish jar

She closes the bottle of root beer

Kyle closes the jar in disappointment

Sally closes the nail polish jar

She closes the tube of ointment

The chef closes the jar of honey

She closes the soda bottle

James closes the fuel cap on his car

Dave closes the peanut butter tub

His mother closes the bottle

Rick closes the bottle of smelly vinegar

Jeff closes the apple juice bottle

She closes the top on the thermos

Megan closes the bottle of cola

Container (Counterclockwise)

The teacher opens the jar of beads

Greg opens the tub of peanut butter

Gwen opens the top of the fuel can

Her husband opens the jar for her

Diane opens the hand-soap pump

Justine excitedly opens the jar

The boy smoothly opens the lotion tub

Zach's opens the plastic water bottle

He opens the child-proof bottle

After a struggle he opens the jar

He opens the cap on his toothpaste tube

She opens an orange juice bottle

She opens the thermos lid

She forcefully opens the mason jar

Jenny opens the almond butter jar

She closes the water bottle at the gym

Non-container (Clockwise)

Ashley winds forward the egg timer

His pencil is dull so he sharpens it

Chad turns the combination lock to right-17

Using the dimmer he turns up the light

He moves forward the hand on the clock

Casey tightens the screw in the deck

Martin tightens the bolt in the frame

She turns up the quiet radio

She winds her watch forward

The driver turns on his engine

Ross exits the highway to visit a restaurant

He tightens the screw in the frame

Ginger screws a light bulb into the lamp

Sebastian tightens a loose screw on his glasses

Jerry screws an audio cord into the TV

Kevin turns on the cars ignition

She sneakily opens the Nutella jar

Non-container (Counterclockwise)

He unscrews the old light bulb

Adrian removes the screw from the faceplate.

The experienced truck driver turns left

Using the dimmer he turns down the lights

She turns down the radio volume

He turns off his cars engine

Jared unscrews the old showerhead from the pipe

Luis unscrews the old light bulb

Alexander turns off the oven

He turns down the loud stereo

She winds back the time on her watch

Jacob unscrews a metal screw

She turns down the car radio volume

Lindsey unscrews the lugnut on her flat tire

Jimmy unscrews the smoke detector

Dan winds back the clocks hour hand

A.3 Critical Stimuli used in Experiment 2

Container (Clockwise)

Outside the gym she fills her water bottle from the fountain and securely closes it as she runs back to practice.

The jar of relish had been left out without a lid so Harvey closes it before refrigerating the jar.

His bottle of root beer is beginning to go flat so he closes the bottle to trap in the bubbles.

Masking his disappointment at the jars contents, Kyle smiles as he grabs the lid and closes the jar gently.

The gold nail polish in the bottle is exposed and drying out, so Pam grabs the top and closes it tightly.

After applying the cream to her scar, she takes the cap and closes the tube of smelly medicine.

The chef pours the sauce into the mason jar and expertly closes it to preserve the Alfredo.

Afraid that the kids will knock over the open cola bottle, she smartly closes the lid to prevent a spill.

The lid doesn't match the jar, so she grabs a new jar and can now close the lid on it.

After filling the car's tank with gas, James grabs the fuel cap and closes it tightly before replacing the pump.

Dave consumed far too much peanut butter so he closes the tub so as not to be tempted by it.

The child is spilling juice from the bottle so his mother grabs it and closes it to his dismay.

Rick pours the vinegar from the bottle and fiddles with the plastic cap before closing the container and putting it away.

Jeff adds the juice concentrate to the bottle and closes the lid tightly before vigorously shaking the container.

She dunks a teabag into the thermos and quickly closes the top so that her tea would stay hot.

Megan decided to save the rest of the cola for later so she closes the bottle and packs it away.

Container (Counterclockwise)

She is craving hazelnut spread, so she takes the jar in her hands and opens it with glee.

Greg grabs the tub of peanut butter for his toast, and grips the lid tightly as he opens it roughly.

The fuel can is empty so Gwen takes it to the pump, grabs the lid and opens the up the top.

The pickle jar lid is on too tightly, so she hand it to her husband, who opens the jar for her.

Diane needs to fill the hand soap bottle, so first she grips the top of the pump and opens the top

Justine is excited to see what is inside the jar and she grabs the lid and opens it forcefully.

With his dry hands, Jeremy grabs the tub of lotion, opens it, and applies the greasy

hand cream.

Zach is thirsty, so when his sister hands him a bottle of water he opens it hastily and drinks it all.

The mason jar is filled with all sorts of goodies, and Lex opens it to sneak a treat before dinner.

The vitamin bottle has a childproof lock, so he pushes down on the cap and then opens it with a twist.

She bought a new tube of toothpaste, and to remove the seal she first opens the cap to access the foil.

The plastic seal around the bottle's lid makes a snapping sound when Shelly opens it during the conference call.

The soup in her thermos is very hot, so to let heat escape she opens it for a minute.

To sanitize the jam jar she grips the metal lid and forcefully opens the container to fill with boiling water.

Jenny wakes and is hungry, so she grabs the almond butter jar and sleepily opens it to eat for breakfast.

The jar lid is sticking so he runs it under hot water and then opens the lid with relative ease.

Non-container (Clockwise)

Kevin hops into the truck, puts the key in the ignition, and turns on the engine, which roars to life.

The blue pencil crayon is dull so he sharpens it before returning it to his teacher.

Chad finally remembers the first number of the combination lock is right 17, and correctly spins the dial to 17.

The light above the painting is on a dimmer knob, so before his guests arrive, Ken turns up the light.

It is springtime daylight savings, so he opens the clock and moves forward the hour hand to 8-o'clock.

The screws in the deck are starting to loosen, so Jenny tightens them until the boards feel more secure.

The bolt in the frame has started to become loose, so Martin grabs his wrench and tightens the bolt.

The radio is too quiet to hear the news so Pat turns up the volume to a more audible level.

Ashley is poaching eggs and grabs her egg timer, winding forward the dial so that it will ring in three minutes.

Kim's watch is running behind, so she checks the time and then winds forward the dial to the correct time.

The driver is waiting at the start line of the race and turns on the engine in anticipation of the starting flag.

Ross sees a sign on the side of the highway and 5 kilometers later he exits the highway to visit the attraction.

To assemble the new bookcase Kailey uses a small wrench to tighten the screws into the wooden frame.

Ginger grabs a new light bulb to put in the empty socket and she screws in the bulb and lights the room.

An arm on his glasses is loose, so Sebastian uses a tiny screwdriver and tightens the small screw gently.

Jerry reaches behind the television and screws in a new audio/visual cord, as the old one was damaged.

Non-container (Counterclockwise)

The clock reads 8 o'clock though it is really only 7, so Dan manually turns back the hour hand.

The outlet's faceplate needs to be replaced so Adrian grabs his screwdriver and removes the screws before removing the plate.

At a four-way stop the truck driver turns on his left indicator light and slowly turns onto Bay Street.

To set the mood Laurence reaches for the dimmer switch and turns down the lights very slowly.

The car radio is too loud to hear her friends so she turns down the volume to a very low level.

He parks the car in the driveway and idles for a minute before finally turning off the loud engine.

Jared wants to install a new showerhead, but first he unscrews the old head from the pipe.

The light bulb in the lamp has burnt out, so Luis unscrews it before replacing the bulb.

The bulb in the slide projector has burnt out, so the professor unscrews the bulb and throws it in the bin.

After pulling the roast out of the oven Kevin grabs the knob on top of the range and turns off the heat.

His parents told him that his stereo was too loud, so he reluctantly turns down the blaring music.

Her watch is running fast so she delicately grips the tiny dial and winds back the time by a few minutes.

Jacob's laptop is overheating so to clean the internal fans he first unscrews the bottom screw to remove the casing.

The ads on the radio are louder than the music so Beth turns down the volume during the commercial break.

In the parking lot Lindsey grabs the last lugnut on her flat tire and unscrews it, throwing it into a dish.

The smoke detector starts beeping in the middle of the night and Jim tiredly unscrews it from the ceiling.

A.4 Critical Stimuli used in Experiment 3a

Note: Slashes represent segment boundaries

Container (Clockwise)

Outside the gym/she fills her water bottle/from the fountain/and securely/closes/it as she/runs back to practice.

The jar of relish/had been left /without a lid on/so Harvey/closes/it before returning/the jar to the fridge

His bottle /of root beer /was beginning to /get flat so he/closes/the bottle to/trap the bubbles in.

Masking his disappointment/at the jars contents, /Kyle smiles as he /grabs the lid and/closes/the jar, while thanking/his secret Santa.

The nail polish in the bottle/is exposed and /drying out so Pam/quickly grabs the top and/closes/it tightly to/preserve the polish.

After applying /the cream/to her scar, she/takes the cap and/closes/the tube/of smelly medicine.

The chef/pours the sauce/into the mason jar/and expertly/closes/it/to preserve the contents.

Afraid that the kids/will knock over/the open bottle of cola/on to the tablecloth, she/closes/the top and puts/the bottle out of reach

Heather places the/lid on top of the jar, /but when she/grabs it and /closes/the lid, it appears to/not fit.

After filling/the car's tank/with gas, James grabs the/fuel cap and/closes/the cap tightly before/hanging up the pump.

After consuming/far too much/peanut butter, Dave/grasps the lid and/closes/the jar/before eating any more.

The child /is spilling juice /from his bottle /so his mother/closes/the bottle, which/causes him to cry.

Rick generously pours /the olive oil/from the bottle, /grips the metal cap and/closes/the bottle/before replacing it on the shelf.

Jeff adds/the juice concentrate/to the bottle/and after he/closes/the bottle lid/he shakes it vigorously.

She dunks/a teabag/into the thermos, /grabs the lid and/closes/the top before/packing it in her bag.

Megan decides to/save the rest/of her cola/for later, so she/closes/the bottle/and puts it in the fridge.

Container (Counterclockwise)

She is craving/hazelnut spread, /so she takes the jar/in her hands and/opens/it gleefully before/digging in with a spoon.

Greg grabs the /tub of peanut butter/for his toast, and grips/the lid tightly as he/opens/it, and stick his knife/into the crunchy spread.

The fuel container/is empty/so Gwen takes it to the pump, /grabs the lid and/opens/the top/so that she can fill it.

The lid on the pickle jar/is on too tightly, /so she hands it to /her husband who/opens/the jar/with relative ease.

Diane needs to/fill the hand soap bottle./so first she grips/the top of the pump and/opens/the top so she may/pour the soap in.

Justine is so excited/to see what is inside/the jar that she grabs /the lid and/opens/it with such force/that the contents fly out.

With his /dry hands/Jeremy grabs the/tub of lotion, /opens/it, and applies the/greasy moisturizer.

Zach is very thirsty, /so when his sister /hands him a bottle/of water he/opens/it hastily/and drinks it all.

The mason jar/is filled with/an assortment/of goodies, and Lex/opens/it to sneak/a taste .

There is a child-proof lock/on the vitamin bottle, /so to access it he pushes down/on the cap before he/opens/the bottle and takes out/a capsule.

The jar lid is sticking/so he puts the top /under the hot water tap/and then forcefully/opens/the metal lid/and dries it off.

The toothpaste cap/is covered in grime, /but she needs to use it/so she pinches the dirty cap and/opens/the top, /to squeeze the paste out.

The plastic seal/around the bottle's lid/makes a satisfying snap/when Shelly/opens/the bottle/and the plastic breaks.

The soup/in her thermos/is very hot/so to let out heat she/opens /the lid/and blows on the soup.

Before she fills/the jar with jam/she must sanitize the inside, /so she grasps the lid and/opens/the jar, /filling it with boiling water.

Jenny is hungry/so she grabs a jar of/almond butter and/gripping the lid she/opens/it, so she can spread/the contents on her toast.

Non-container (Clockwise)

Kevin needs/milk from the store/so he hops in his car/and puts the key in the ignition, /turning it /on as he buckles/his seatbelt.

The blue /pencil crayon/is dull/so he/sharpens/it for/the teacher.

Chad finally remembers/the first number of the/combo lock is right 17/and correctly/spins/the dial/to number 17 .

Excited to try/the wine./Andrew grabs/the corkscrew and/screws/the metal/into the cork.

She reaches/for the/brass key/puts it in the lock, and/unlocks/the door/for her guests.

On the outside deck/the screws in the boards/are starting to come loose/so Jenny/tightens/them so that they do not/stick out.

The radio/is too quiet/to hear the news/so Pat/turns up/the volume/to an audible level.

The chef/clamps the can opener/down on the lip /of the coconut milk can and/twists/the handle until/the lid comes free.

As per/the instructions./Ashley grabs her/egg timer and/winds forward/the dial to the/three-minute mark.

Kim's watch /is running behind/so she checks the/correct time, and then/winds forward/her watch/to match.

Sally calls the operator/on her /grandmother's old/rotary dial phone by/dialing/zero and waiting/to connect.

His car key fob/is broken/so to get inside he/inserts his key in the door and/turns/it manually/until the lock pops.

To assemble /the bookcase/Kailey uses/a little wrench to/tighten/the screw into/the wooden frame.

Ginger grabs/a new light bulb/to put in /the empty socket and/screws/it in until the light/flickers on.

An arm on/his glasses is loose/so using a /tiny screwdriver, Sebastian /tightens/the screw/very carefully.

Jerry reaches behind/the television/to install the/audio/visual cord and/screws in/the cable/to the box.

Non-container (Counterclockwise)

The clock reads 8 o'clock/though it is really/only 7 , so Dan/removes the clock cover and/winds back/the hour hand/by one hour.

The outlet's faceplate/needs to be replaced/so Adrian grabs/his screwdriver and/removes/the screws/before pulling the plate off.

At a four-way stop/the truck driver/turns on his/left indicator light and/turns/onto/Bay Street.

To set the mood/Laurence reaches/for the /dimmer switch and/turns down/the lights/very slowly.

The car radio/is too loud/to hear her friends/so she/turns down/the volume/to a very low level.

His friend was taking /a long time to come out/of her house, /so rather than idle his engine he/turns off/his car/and heads towards the house.

Jared wants to/install a new /showerhead, /but first he /unscrews/the old head/from the pipe.

The light /burnt out/so Luis/has to/unscrew/the old bulb/and replace it.

Martin keeps/burning the soup/so Julia reduces/the heat by/turning/the knob/to a lower setting.

After pulling the roast/out of the oven, /Kevin uses a knob/on top of the range to/turn off/the heat/before eating.

He fumbles/for the switch/under the lampshade, /finally finding it and/turning off/the lamp /to save electricity.

Her pocketwatch/is a few minutes fast/so she delicately grips/the tiny dial and/turns back/the time/by 3 minutes.

Jacob's laptop/is overheating/so to clean it/he first/unscrews/the bottom screw/and pries the base off.

The ads on the radio/are so much louder/than the music/so Beth/turns down/the volume/at the commercial break.

With her car/hoisted up on the jack, /Lucy grabs hold /of a lug nut and/loosens/it, throwing the/nut into a dish.

The smoke detector/starts beeping/in the middle of the night/and Jim tiredly/unscrews/it to replace/the batteries.

A.5 Critical Stimuli used in Experiment 3b

Note: Slashes represent segment boundaries.

Container (Clockwise)

Outside the/gym she /fills her/water bottle /from the/fountain/and securely/closes/it as she/runs back/to practice.

The jar/of relish/had been/left out/without/a lid/so Harvey/closes/it before/refrigerating/the jar.

His bottle/of root/beer is/beginning to/go flat/so he/closes/the bottle/to/trap in/the bubbles.

Masking his/disappointment/at the/jars contents,/Kyle smiles/as he/grabs the/lid and/closes/the jar/gently.

The gold/nail polish/in the/bottle is/exposed and/drying out, /so Pam/grabs the/top and/closes/it tightly.

After applying/the cream/to her/scar, she/takes the/cap and/closes/the tube/of/smelly/medicine.

The chef/pours the/sauce/into the/mason jar/and expertly/closes/it to/preserve/the/alfredo.

Afraid that/the kids/will knock/over the/open cola/bottle, /she smartly/closes/the lid/to prevent/a spill.

The lid/doesn't match/the jar, /so she/grabs a/new jar/and can/now/close/the lid/on it.

After filling/the car's/tank with/gas, James/grabs the/fuel cap/and/closes/it tightly/before replacing/the pump.

Dave consumed/far too/much/peanut butter/so he/closes/the tub/so as/not to/be tempted/by it.

The child/is spilling/juice from/the bottle/so his/mother grabs/it and/closes/it to/his/dismay.

Rick pours/the vinegar/from the/bottle and/fiddles with/the plastic/cap before/closing/the container/and putting/it away.

Jeff adds/the juice/concentrate/to the/bottle and/closes/the lid/tightly before/vigorously/shaking/the container.

She dunks/a teabag/into the/thermos/and quickly/closes/the top/so that/her tea/would stay/hot.

Megan decided/to save/the rest/of the/cola/for later/so she/closes/the bottle/and packs/it away.

Container (Counterclockwise)

She is/craving/hazelnut spread,/so she/takes the/jar/in her/hands and/opens/it with/glee.

Greg grabs/the/tub of/peanut butter/for his/toast, and/grips the/lid tightly/as he/opens/it roughly.

The fuel/can/is empty/so Gwen takes/it to/the pump, /grabs the/lid and/opens/the up/the top.

The pickle/jar lid/is on/too tightly,/so she/hand it/to her/husband, who/opens/the jar/for her.

Diane needs/to fill/the hand/soap bottle,/so first/she grips/the top/of the /pump and/opens/the top

Justine is/excited/to see/what is /inside/the jar/and she/grabs the/lid and/opens/it forcefully.

With his/dry hands,/Jeremy grabs/the tub/of lotion,/opens/it, and/applies the/greasy/hand/cream.

Zach is/thirsty, so/when his/sister hands/him a/bottle of/water he/opens/it hastily/and drinks/it all.

The mason/jar is/filled with/all sorts/of goodies, /and Lex/opens/it to/sneak a/treat before/dinner.

The vitamin/bottle has/a child-proof/lock, so/he pushes/down on/the cap/and then/opens/it with/a twist.

The jar/lid is/sticking so/he runs/it under/hot water/and then/opens/the lid/with relative/ease.

She bought/a new/tube of/toothpaste, and/to remove/the seal/she first/opens/the cap/to access/the foil.

The plastic/seal around/the bottle's/lid makes/a snapping/sound when/Shelly/opens/it during/the conference/call.

The soup/in her/thermos/is very/hot, so /to let/heat escape/she/opens/it for/a minute.

To sanitize/the jam/jar she/grips the/metal lid/and forcefully /opens/the container/to fill/with boiling/water.

Jenny wakes/and is/hungry, so/she grabs/the almond/butter jar/and sleepily/opens/it to/eat for/breakfast.

Non-container (Clockwise)

Ashley is/poaching eggs/and grabs/her egg/timer, /winding forward/the dial/so that/it will/ring in. /three minutes.

Kim's watch/is running/behind,/so she/checks the/time and/then/winds forward/the dial/to the/correct time.

The driver/is waiting/at the/start line/of the/race and/turns on/the engine/in anticipation/of the/starting flag.

Ross sees/a sign/on the/side of/the highway/and 5/kilometers later/he exits/the highway/to visit/the attraction.

To assemble/the new/bookcase/Kailey uses/a small/wrench to/tighten/the screws/into the/wooden/frame.

Ginger grabs/a new/light bulb/to put/in the/empty socket/and she/screws in/the bulb/and lights/the room.

An arm/on his/glasses is/loose, so/Sebastian uses/a tiny/screwdriver and/tightens/the small/screw/gently.

Jerry reaches/behind the/television/and/screws in/a new/audio/visual/cord, as/the old/one was/damaged.

Kevin/hops /into the/truck, puts/the key/in the/ignition, and/turns on/the engine/which roars/to life.

The blue/pencil/crayon/is dull/so he/sharpens/it before/returning/it to/his/teacher.

Chad finally/remembers the /first number/of the/combination lock/is right/17, and /correctly/spins/the dial/to 17.

The light/above/the painting/is on/a dimmer/knob,/so before/his guests/arrive, Ken/turns up/the light.

It is/springtime/daylight/savings,/so he/opens the/clock and/moves forward/the hour/hand to/8-o'clock.

The screws/in the/deck/are starting/to loosen,/so Jenny/tightens/them until/the boards/feel more/secure.

The bolt/in the/frame/has started/to become/loose, so/Martin/grabs his/wrench and/tightens/the bolt.

The radio/is too/quiet/to hear/the news/so Pat/turns up/the volume/to a/more audible/level.

Non-container (Counterclockwise)

The bulb/in the/slide projector/has burnt/out, so/the professor/unscrews/the bulb/and throws/it in/the bin.

After pulling/the roast/out of/the oven/Kevin grabs/the knob/on top/of the/range and/turns off/the heat.

His parents/told him/that his/stereo/was too/loud, so/he/reluctantly/turns down/the blaring/music.

Her watch/is running/fast so/she delicately/grips the/tiny dial/and/winds back/the time/by a/few minutes.

Jacob's laptop/is overheating/so to/clean the/internal fans/he first/unscrews/the bottom/screw/to remove/the casing.

The ads/on the/radio are/louder than/the music/so Beth/turns down/the volume/during the/commercial/break.

In the/parking lot/Lindsey grabs/the last/lugnut on/her flat/tire and/unscrews/it, throwing/it into/a dish.

The smoke/detector/starts beeping/in the/middle of/the night/and Jim/tiredly/unscrews/it from/the ceiling.

The clock/reads 8/o'clock/though it/is really/only 7, /so Dan/manually/turns back/the hour/hand.

The outlet's/faceplate/needs to/be replaced/so Adrian/grabs his/screwdriver and/removes/the screws/before removing/the plate.

At a/four-way/stop the/truck driver/turns on/his left/indicator light/and slowly/turns/onto/Bay Street.

To set/the mood/Laurence/reaches/for the/dimmer/switch and/turns down/the lights/very/slowly.

The car/radio is/too loud/to hear/her friends/so she/turns down/the volume/to a/very low/level.

He parks/the car/in the/driveway/and idles/for a/minute before/finally/turning off/the loud/engine.

Jared wants/to install/a new/showerhead, /but first/he/unscrews/the old/head/from the/pipe.

The light/bulb/in the/lamp has/burnt/out, so/Luis/unscrews/it before/replacing/the bulb.