MODELING THE MECHANISMS OF VERB BIAS LEARNING

BY

AMANDA KELLEY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Gary S. Dell, Chair
Professor Cynthia L. Fisher
Professor Kara D. Federmeier
Assistant Professor Jon Willits
Assistant Professor Jessica L. Montag

ABSTRACT

During language production, speakers acquire statistical information about linguistic units. One example of this occurs in verb bias learning, in which speakers update statistics about how likely a particular verb is to occur in a specific syntactic structure. This dissertation explores the mechanisms that support verb bias learning using a combination of empirical work and cognitive models of empirical results. Chapter 2 implements two cognitive models that reproduce previous empirical results. The models in this section behave similarly to humans when asked to learn and unlearn rules, and when they learn expected and unexpected verb-structure combinations. Chapters 3 and 4 detail two experimental investigations into how learning a new bias for one verb can transfer to another, semantically-related verb. Chapter 3 details the initial investigation, and shows that verb bias learning can transfer to similar dative verbs, but that transitive verbs show neither training nor transfer. Chapter 4 explains two follow-up norming studies to select new verb pairs, and then a behavioral replication of the study in Chapter 3. This new study shows no transfer for the new set of dative verbs, and replicates the finding of no training and no transfer for the transitive verbs. Chapter 5 uses a cognitive model to generate the results of the two dative transfer studies, and shows that the finding of transfer in Chapter 3 but not in Chapter 4 can be explained by a difference in how unexpected a structure is for each verb. Chapter 6 models the transitive results found in Chapters 3 and 4, and shows that a predictive object-first cue effectively blocks verb bias learning, consistent with the possibility that transitive learning is blocked by these or related cues. Finally, Chapter 7 addresses findings from Lin (2020) and Thothathiri and Braiuca (2021), which explore how humans switch from using verbs to predict sentence structure to using other cues. However, the models struggle to replicate these results, which suggests a need for a more complex model. This dissertation shows

that many of the findings in the verb bias learning literature can be accounted for using the proposed cognitive model, and contributes new findings that also fit within this framework. Additionally, it lays out specific reasons why the model fails to account for cue-switching results, and how improvements to the model could guide further empirical research.

ACKNOWLEDGEMENTS

Over the last seven years, I have been extremely grateful for the communities that I have been a part of. The following are some of the people who have made up parts of the gestalt.

First, I would like to thank my advisor, Gary Dell. I have had the great privilege of learning from him about language production and cognitive modeling, intellectual curiosity, and how to be a good colleague. I am grateful for his excitement about my work when I couldn't summon any myself, and for the many, many emails he answered. This dissertation would not exist without his support and wisdom.

I would also like to thank my committee: Cynthia Fisher, Kara Federmeier, Jon Willits, and Jessica Montag. I especially thank Cindy for her thoughtful questions, and for serving on every committee that I have convened since my first year. I also thank Kara for her foresight and attention to detail not only in her service on my committee, but in the process of developing and teaching 224.

I also want to thank my fellow graduate students in the Psychology department. In particular, I want to acknowledge Jessica Siler and Melinh K. Lai, who were always there when I needed a road brother. I also need to thank the members of the Language Processing Brown Bag and the Cognitive division more generally. For those who finished and those who didn't, it wouldn't have been the same without you.

I would not have even started without my parents, whose support and love have been there for me since finishing a science project meant gluing a bunch of construction paper onto a trifold. I am also indebted to my sister, Erin, who is a fearless voice of reason. I also need to thank Katherine Liu, who has been a source of encouragement since all the way back in high school. Finally, I need to thank the Wszaleks, who have offered their support and some much-needed dinners throughout this process.

Last but most, I thank Andy. He lived in the thick of it with grace and optimism, and I cannot imagine how my life would look without him.

TABLE OF CONTENTS

CHAPTER 1: REVIEW OF VERB BIAS LEARNING

Language is frequently ambiguous and developing strategies to reduce ambiguity is one goal of language learning. However, language also contains statistical regularities that allow speakers to predict what sorts of linguistic events are likely to occur, and these regularities exist at all levels. For example, recurring groups of syllables are more likely to be words, and humans are sensitive to the probability that two syllables will occur together (Saffran, Aslin, and Newport, 1996). Importantly, these regularities are not static. Instead, they are frequently updated, even in a first language. For example, English speakers can learn that the phoneme /f/ is constrained to occur only in the first position of a syllable, e.g.,"fes", even though this is not true in English (Dell, Reed, Adams, and Meyer, 2000). This kind of learning allows language knowledge to be flexible, while also taking advantage of a lifetime's worth of experience with language.

One domain where flexible language knowledge is exhibited is in the probabilistic relationship between verbs and sentence structures, also called a verb bias. A verb is said to be biased when it tends to occur in a specific sentence structure, although it may have the ability to occur in others. For example, the verb "accept" tends to occur in direct object sentences, such as "The usher *accepted* the ticket" (Garnsey, Pearlmutter, Myers, and Lotocky, 1997). By contrast, the verb "admit" tends to occur in sentential complement sentences, such as "The usher *admitted* that he had lost the ticket." Reading or producing verbs in a structure that they are biased toward is easier than in structures that they are not biased toward (Garnsey et al. 1997, Ferreira and Schotter, 2013). For example, reading a sentential complement structure with "accept" as its verb will tend to be more difficult than reading the same structure with "admit" (Garnsey et al. 1997). Consequently, verb biases can reduce ambiguity in comprehension and ease language production. Verb biases are also learned and updated. Adults and children can learn new verb biases for familiar verbs (Coyle and Kaschak, 2008; Lin and Fisher, 2017; Qi, Yuan, and Fisher, 2011), and adults can acquire and use verb biases for novel verbs as well (Wonnacott, Newport, and Tanenhaus, 2008). Finally, the patterns acquired through this learning process are not simple associations between single verbs and structures. Instead, they allow learning of a verb bias in some situations and learning about the general tendency of large groups of verbs in others (Wonnacott et al. 2008, Thothathiri and Braiuca, 2021). Verb bias is not only a good example of how linguistic regularities are used, but also a good domain to understand how this learning takes place.

Although verb biases have been well-studied as a component of human language use, the processes that govern updating are a newer area of study. Consequently, the ways in which verb bias learning is like other kinds of statistical language learning are not well understood, and the mechanistic details of how that learning occurs are underspecified. The goal of this review is to first establish what role verb biases play in language comprehension and production throughout the lifespan. Secondly, it will explain what is known about how verb biases are learned, including studies that directly document this learning. Finally, this review will demonstrate how current research could be enhanced by the addition of a computational model of verb bias learning.

1.1 VERB BIAS EFFECTS IN LANGUAGE COMPREHENSION AND PRODUCTION

Verb bias affects both the process of understanding language, and the process of producing it. These effects are shown by both children and adults. Additionally, these effects do not seem to have a strong developmental gradient. Rather, adults and children both show similar effects of verb biases in comprehension and production. Comprehension studies of learners of English as a second language suggest that verb biases can also be learned in a second language, and similar studies with people with a language deficit called aphasia find that verb bias effects remain even when the language system has been damaged. In short, verb bias is a core part of linguistic knowledge for speakers of English, and this knowledge is demonstrated in both language comprehension and production.

1.1.1 Comprehension

Verb biases were originally studied as a factor in adult language comprehension. Although one early study questioned whether verb biases played a role in comprehension (Ferreira and Henderson, 1990), currently verb biases are known to be one of many constraints that guide language comprehension (e.g., Garnsey, Pearlmutter, Myers, and Lotocky, 1997). This is true not only for typical adult native speakers, but also for children (e.g., Snedeker and Trueswell, 2004), people who have learned English as a second language (e.g., Dussias and Cramer Scaltz, 2008), and people with aphasia (e.g., Dede, 2013). For all these groups, it is generally true that reading or listening to sentences where a verb's bias matches the sentence's structure is facilitated. However, these groups differ in how they are able use other cues to sentence structure alongside verb biases.

For typical adults who speak English as their first language, verb biases begin to indicate sentence structure relatively early in processing. Verb biases were first studied to investigate how comprehenders process the syntactic structure of a sentence. One theoretical claim was that more complex structures should always be read more slowly than simple sentences (e.g., Frazier and Rayner, 1982). Consequently, this perspective predicts that an association between a specific verb and a specific syntactic structure should not influence reading (Ferreira and Henderson, 1990). However, other proposals suggested that comprehenders use lexical information to anticipate sentence structure, even in cases where the predicted structure would not be simple (see Mitchell, 1989, for a contemporary review). To study these predictions, multiple studies compared sentential complement sentences with main verbs that were either biased toward or away from a sentential complement, like those below (bias norms from Garnsey, Pearlmutter, Myers, and Lotocky, 1997):

a. The witness admitted (that) the fraud had started several years earlier. (Sentential Complement bias)

b. The witness confirmed (that) the statements he had made before were false. (Direct Object bias)

Overall, verb biases seem to immediately facilitate reading sentences where the verb and structure match (Trueswell, Tanenhaus, and Kello, 1993; Garnsey et al., 1997). Whether the second noun in the sentence is a plausible direct object – in other words, whether "The witness confirmed the statements" is also possible sentence – does not affect reading times when a verb is strongly biased toward a sentential complement (Garnsey et al. 1997). However, Garnsey and colleagues note that comprehension is more difficult when the structures suggested by the second noun and the verb's bias conflict, suggesting that these two cues are both used up until the point at which the sentence is disambiguated. In conclusion, Garnsey and colleagues report that verb biases played an early and consistent role in sentence comprehension, with a smaller role of plausibility reserved for sentences with particularly ambiguous completion. These results also show that verbs exist within an ecosystem of other information that also helps with disambiguating sentence structure.

By contrast, although children can successfully use verb biases during language comprehension, they seem to have more difficulty weighing verb biases against other cues to sentence structure. Building on findings that children have more difficulty both reanalyzing temporarily ambiguous “garden path” sentences and using contextual information to constrain which structures to consider (e.g., Trueswell, Sekerina, Hill, and Logrip, 1999), Snedeker and Trueswell (2004) investigated how children use contextual information and verb biases to comprehend ambiguous sentences. Snedeker and Trueswell used sentences like those in c:

c. Feel the frog with the feather.

These sentences have two potential interpretations; the frog could be in possession of the feather (a modifier interpretation), or the feather could be used to touch the frog (an instrument interpretation). Snedeker and Trueswell used verbs that varied in their preferences for each interpretation. For example, a verb like “feel” is about equally likely to occur with either a modifier or an instrument interpretation. However, a verb like “choose” tends to result in a modifier interpretation, while a verb like “poke” will tend to create an instrument interpretation. In order to understand how these different biases influenced how participants interpreted sentences, participants were presented with a display of objects that included a potential instrument (a feather), a target (a frog with a feather), and either the same animal (a frog with another object) or a different animal. Snedeker and Trueswell found that both five-year-old children and adults were able to use verb biases to aid language comprehension, but that adults were more able to use referential information as well. Kidd, Stewart, and Serratrice (2011) examined sentences like those in Snedeker and Trueswell (2004), but with instrument-biased verbs that were paired with either a plausible or implausible instrument. For example:

d. Cut the cake with the candle (Implausible)

Kidd et al. found that adults were able to somewhat overcome the bias for the verb “cut” for an instrument (d.) when the potential instrument is implausible (candle). Their five-year-old participant group, though, seemed to be particularly affected by the verb’s bias toward the (here, implausible) instrument interpretation. Intriguingly, these results suggest that verb biases are available to children relatively early in development. However, the use of other cues, such as the implausibility of cutting a cake with a candle, develops later.

Second language speakers of English also use verb biases during language comprehension but may not always use them in a native-like way. Dussias and Cramer Scaltz (2008) examined a group of bilingual Spanish-English speakers and a group of monolingual English speakers who read sentential complement sentences (similar to examples a and b above). Dussias and Cramer Scaltz found that bilingual Spanish-English speakers used verb bias information in much the same way as monolingual English speakers did, provided that they had fully acquired English verb biases and were not importing conflicting ones from Spanish. For languages that are less English-like, the ability to use cues like verb bias seems to be somewhat less natural for second-language speakers. For example, Lee, Lu, and Garnsey (2013) presented sentential complement sentences to a group of bilingual Korean-English speakers and an English-speaking control, and found that lower-proficiency Korean-English bilinguals were less able to use verb biases in a native-like way when the complementizer “that” was not present. Lee and colleagues suggest that this reflects difficulty using a verb bias without additional support from the complementizer, which may reflect difficulty on the part of the lower-proficiency speakers in learning a parsing strategy that is not useful in Korean, where verbs are frequently at the end of a sentence.

Similar results were found by Qian, Lee, and Lu (2019), who found that bilingual Mandarin-English speakers also had difficulty using the presence of a “that” in the complement clause and the bias of a particular verb together as clues to sentence structure. However, the Mandarin-English bilinguals did not behave differently at higher and lower levels of English proficiency. Together with evidence that the Mandarin-English bilinguals seem to have more detailed verb bias representations, Qian et al. suggest that this may be because the structure of Mandarin tends to place the verb in the middle of sentences, which is more similar to English than the typical sentence structure of Korean. Consequently, Mandarin-English bilinguals may gain an advantage from transfer from their first to their second language. Finally, Anible, Twitchell, Waters, Dussias, Piñar, and Morford (2015) come to a similar conclusion concerning deaf ASL-English bilinguals, who use verb bias information in some situations, but also generally prefer the more-complex sentential complement structure compared to the simpler direct object sentence structure. Anible et al. suggest that this may be the result of parsing strategies that are tailored for visual languages like ASL rather than spoken language like English. Consequently, it appears that for second languages with significant syntactic overlap with the native language (e.g., Spanish for English) and for high-proficiency second-language speakers, verb biases are an inherent part of language comprehension. For lower-proficiency speakers or speakers whose languages have less overlap with English, it is more difficult to adapt their parsing strategies, and consequently there are qualitative differences in their ability to use these cues.

Finally, several studies have addressed whether people who have impaired language abilities because of brain damage (aphasia) continue to use verb biases during language comprehension (see Gahl and Menn, 2016, for review). Although the results are noisier than for healthy adults, Gahl and Menn argue that studies of people with aphasia generally find that they rely heavily on verb biases, perhaps even more extensively than healthy adults. This can be seen particularly clearly in studies like Dede (2013), which compared the performance of people with aphasia to healthy controls using a subset of sentences from Garnsey, Pearlmutter, Myers and Lotocky (1997). While people with aphasia did show evidence that verb biases facilitated their reading, they did not show the sensitivity to the presence or absence of "that" or the plausibility of the direct objects that healthy controls did (Dede, 2013). In conclusion, verb biases are not only used to comprehend specific types of sentences; instead, they are part of a general comprehension strategy that is used by both healthy adult language users and other groups, such as people with aphasia.

Overall, verb biases are one of many constraints that are used during language comprehension. They are available relatively early in development (e.g., Snedeker and Trueswell, 2004), and are also used by groups like second language speakers of English (e.g., Lee, Lu, and Garnsey, 2013) and people with aphasia (e.g., Dede 2013), whose knowledge could be incomplete or damaged. In general, it is not access to verb biases that causes differences in reading strategies; rather it is the ability to balance different cues against each other. In sentential complement sentences, typical adult native English speakers can use verb biases, plausibility of the second noun, and the presence of *that* to understand sentences (Garnsey, Pearlmutter, Myers, and Lotocky, 1997). Second language speakers of English cannot universally use these same cues (Lee, Lu, and Garnsey, 2013; Qian, Lee, and Lu, 2020), and children seem to struggle with processes that allow them to reanalyze sentences that do not conform to a verb's bias (Kidd, Stewart, and Serratrice, 2011). Verb biases are an important cue in language comprehension and may also be generally more available or accessible than other cues to sentence structure.

1.1.2 Production

Given the clear role verb biases play in language comprehension, it naturally follows to ask how they influence language production. For adults, the effects seem to be twofold: verb biases influence which structures are produced in the first place, and the repeated pairing of verbs and structures changes articulatory aspects of how these structures are produced. It is easier to produce sentences with structures that match the bias of the main verb (Ferreira and Schotter, 2013). Speakers also show signs of articulatory reduction when producing sentences that use structures that match the bias of the main verb, suggesting that these learned probabilities affect how sentences are produced (e.g., Gahl and Garnsey, 2004). Finally, when verb bias effects are observed in children, effects on production can be seen in three-year-olds (Kidd, Lieven and Tomasello, 2006; Peter, Chang, Pine, Blything, and Rowland, 2015). Jointly, these sets of findings show that verb biases facilitate language production for common verb-structure pairings, and that this facilitation is present throughout the lifespan.

Verb bias statistics clearly influence language production in part by encouraging the production of common verb-structure pairings. One line of evidence comes from studies like Garnsey, Pearlmutter, Myers and Lotocky (1997), which used a sentence completion task to determine the biases of verbs included in their study. While offline tasks like these are not the most sensitive way to measure production, they do suggest that the same statistics affect comprehension and production. Bernolet and Hartsuiker (2010) approached this problem by examining the effect of verb biases on syntactic priming. Syntactic priming describes facilitated processing for a syntactic structure after it has been seen recently (Bock, 1986). Bernolet and Hartsuiker found that the strongest priming was caused by sentences whose verbs' biases did not match the sentence structure. For example, if a verb was strongly biased toward the prepositional dative (giving something to someone), seeing that verb in a double object sentence (giving someone something) created stronger priming than verbs that were already biased toward the double object dative. Ferreira and Schotter (2013) also address why verbs tend to be produced in their preferred structures, using the observation that including an optional “that” in sentential complement sentences can be an index of greater production effort. Ferreira and Schotter found that overall, “that” was produced more often with verbs biased against a sentential-complement continuation. For example, the verb “accept” tends not be used as a sentence complement. But if a speaker does use it with that structure, e.g., then “that” is more likely to be included. Additionally, producing or not producing “that” was not conditioned on the potential for a sentence to be ambiguous, but instead on how unlikely that verb was to be produced with a sentential complement. In other words, Ferreira and Schotter (2013) suggest that a verb bias toward a sentential complement reduces production difficulty for that sentence, while using a verb with any structure that is not its preferred structure will result in greater production difficulty. Although the results of Bernolet and Hartsuiker (2010) and Ferreira and Schotter (2013) use different methodologies, both suggest that production is sensitive to co-occurrence of verbs and sentence structures. Consequently, these studies suggest not only that production tracks verb bias statistics, but also that the most common completion for a verb seems to be privileged.

Further evidence that verb biases are best characterized as facilitation toward a specific structure comes from studies of the articulatory properties of sentences. Gahl and Garnsey (2004) used markers like word duration and reductions like the deletion of t’s or d’s at the end of words like “and” to study verb bias effects while producing direct object or sentential complement sentences. Gahl and Garnsey found that /t-d/-deletion increased when a verb’s bias matched the sentence structure, and that duration of features like clause boundaries was longer when the structure was unlikely given the verb. A similar effect was found in Gahl, Garnsey, Fisher and Matzen (2006), which examined sentences with verbs that could either be transitive or intransitive. Tily, Gahl, Arnon, Snider, Kothari, and Bresnan (2009) extended this method to spontaneous rather than recited speech. Tily and colleagues did not find a significant effect of verb bias, but suggest that this is because they included information like givenness and animacy in their models, which may have accounted for variance in structure choice that is normally accounted for by verb biases. Collectively, these studies suggest that production is sensitive to the probability that a word will occur in a particular sentence structure. Whether this probability is best explained by statistical co-occurrences or correlated factors that influence production remains an open question, but it is possible to say that verb biases can account for this data when other fine-grained predictors are not included.

Children also show evidence that verb biases influence their language production. Peter, Chang, Pine, Blything, and Rowland (2015) studied the performance of a group of three- and four-year-olds, a group of five- and six-year-olds, and a group of adults on a syntactic priming task. They found that both groups of children were more likely to produce the structure that a target verb was biased toward, and that they were more likely to repeat a structure when they were primed with a verb that was biased against that structure. In other words, Peter et al. found that children showed sensitivity to verb biases in language production in ways that are similar to the adults in their study and to the adult participants in Bernolet and Hartsuiker (2010). Kidd, Lieven, and Tomasello (2006) also suggest that young children have fine-grained knowledge of how frequently verbs appear in particular structures. Using a sentence repetition task, Kidd et al. found that a group of children ranging from 2 years, 10 months to 4 years, 2 months were more likely to correctly repeat sentences containing verbs that are likely to appear in that sentence structure, and that they were more likely to substitute highly biased verbs when they repeated a sentence incorrectly. Kidd, Lieven, and Tomasello (2010) followed up on this work using a priming task with groups of four-year-olds and six-year-olds, who were primed with sentences that contained verbs that were either likely or unlikely given the sentence structure. Kidd et al. again found that children were more successful when they were asked to recall a sentence with a likely verb, and that children tended to substitute likely verbs for unlikely verbs when they recalled incorrectly. Consequently, these studies suggest that children already show knowledge of verb biases in language production almost as early as their production can be reliably measured. In tandem with the comprehension literature, these results suggest that even young children use their knowledge of verb biases during language production.

The effects of verb biases in production parallel their effects in language comprehension. Saying sentences where the structure matches the verb bias is easier than when the structure does not match the bias (Ferreira and Schotter, 2013). Additionally, verb biases change the articulation when sentence structure and verb bias match (e.g., Gahl and Garnsey, 2004). Together, these results show that adults have learned which structures and verbs commonly occur together, and that this knowledge affects both what they choose to produce and how they choose to produce it. Finally, verb biases do not only affect adult production; they also affect production at least by age three (Peter, Chang, Pine, Blything, and Rowland, 2015; Kidd, Lieven, and Tomasello, 2006). Verb bias effects in production begin early and persist into adulthood, forming the basis for production effects documented in adults.

1.2 LEARNING VERB BIASES

Although verb biases clearly play an important role in language comprehension and production, the question of where they come from encompasses literatures that take very different approaches to solving this problem. One way to address this is to study children, whose behavior demonstrates how and when verb biases are acquired (e.g., Snedeker and Trueswell, 2004; Peter, Chang, Pine, Blything, and Rowland, 2015). The other is to teach participants a new verb bias in the lab, which hinges on the prediction that verb biases naturally originate from consistent verb-structure pairings and that this process can be replicated in a controlled environment (e.g., Coyle and Kaschak, 2008). Many of these studies focus on understanding how distributional learning can account for verb bias learning. Finally, a smaller group examines the potentially parallel contribution of meaning. Together, this body of work is discovering how verb biases develop, and where the knowledge to build them comes from.

Children clearly show effects of verb biases in their language comprehension and production. As reviewed above, they use verb biases to predict upcoming structures during language comprehension (Snedeker and Trueswell, 2004; Kidd, Stewart, and Serratrice, 2011). They also show effects of verb biases in production; for example, showing stronger priming when a verb is paired with a structure it does not prefer (Peter, Chang, Pine, Blything, and Rowland, 2015). Because the available evidence suggests that verb bias knowledge is adult-like in three- and four-year-olds, it is likely that this knowledge was acquired even earlier (e.g., Fisher, Jin, and Scott, 2020). Demonstrating how this knowledge is acquired involves both identifying an early learning mechanism that is capable of learning statistical information, and showing that children can learn about verbs from an early age.

There is abundant evidence that babies begin to learn about environmental statistics from a very early age. This is thought to be done by distributional learning, a general learning mechanism that allows children to keep track of statistical regularities that they experience (see Gomez and Gerken (2000) for further review). In particular, distributional learning can help children track repeating sequential stimuli, which includes information like the words that tend to immediately surround verbs (see Romberg and Saffran (2010) for further discussion). There is also evidence that children use their input to learn about the behavior of verbs. Children tend to use individual verbs in the kinds of sentences that their caregivers use those verbs in (Theakston, Lieven, Pine, and Rowland, 2001), and use the distributions of multiple types of sentences to narrow down the structural biases of individual verbs (e.g., Twomey, Chang, and Ambridge, 2014). Although children are clearly able to use the language statistics that are available to them to begin learning about verb behavior, experimental studies are also necessary to understand what aspects of the input are most helpful to children.

Unlike studies of more naturalistic input, experimental studies can carefully control the kinds of information a child receives about a verb. These studies show that children as young as two can use distributional cues to learn about the behavior of verbs. For example, Scott and Fisher (2009) investigated the case of learning distributions for the difference between two classes of verbs: causal verbs that highlight changes to an object, and unspecified-object verbs that highlight the action being done. Both types of verbs can both occur in transitive sentences, such as g., but alternate to a different intransitive structure as in h. and i:

g. Anne broke/dusted the lamp.

h. The lamp broke. (Causal)

i. Anne dusted. (Unspecified-Object)

Scott and Fisher found that verbs like “dust” and “break” vary on a number of distributional parameters in child-directed speech, including whether their intransitive form takes an animate agent, which could be tracked to learn the difference between these verb classes. In a separate experiment, 28-month-old children listened to a dialogue that used those distributional parameters to indicate which class a novel verb belonged to. When presented with an ambiguous transitive sentence and either a causal or unspecified-object video, Scott and Fisher found that children tended to look at the video that was congruent with the distribution they had previously heard used with that verb. This suggests that children track relevant parameters that allow them to assign verbs to a correct class, including which syntactic environments they have seen a verb in. Further work also suggests that two-year-olds apply this same kind of learning to other verb classes, such as transitive and intransitive verbs (Yuan and Fisher, 2009), and that children may use other cues such as discourse structure to further organize the input they receive to learn about verb behavior (see Fisher, Jin, and Scott 2020 for a comprehensive review of how children learn verb distributions).

Although verb class information is very similar to verb bias information, they are not identical. Verb biases involve the probabilistic pairing of one verb and one structure, while verb classes refer to the general behavior of a group of similar verbs. However, experimental work with slightly older children does show that children can use distributional learning to acquire new verb biases. Qi, Yuan, and Fisher (2011) trained five-year-old children using dialogues that either encouraged an instrument or a modifier interpretation of a verb. Like in Snedeker and Trueswell (2004), children were presented with sentences like “Feel the frog with the feather”, where the feather could be either an instrument used to feel a frog, or a modifier used to describe a specific frog. Qi et al. (2011) found that with relatively minimal training, children began to look toward the interpretation of that verb that they were newly biased toward, suggesting that they had changed their biases for that verb based on recent experience. Lin and Fisher (2017) extended these findings with four- and five-year-olds, finding that children not only learned new verb biases in production, but also that training children toward an unlikely structure given the verb's pre-existing bias increased the size of the training effect, which was further modulated by how common a particular syntactic structure is. This result can be characterized as evidence that children track the contexts of individual lexical items as well as the relative frequency of the syntactic structures themselves. Lin and Fisher also note that this behavior can be characterized as surprisal, or a greater response to unexpected stimuli, and is consequently a hallmark of implicit learning. Collectively, these studies show that children learn verb biases from their input, using a distributional learning mechanism.

Adults also retain the ability to adjust verb biases in their first language. Coyle and Kaschak (2008) presented adult native speakers of English with a series of dative sentences that always contained only one verb. For example, the verb "send" might only occur in double object dative sentences (e.g., I sent her a package), and the verb "hand" might only appear in prepositional dative sentences (e.g., I sent a package to her). After this training, Coyle and Kaschak found that their participants were more likely to produce verbs in the structure they had just been experienced in, suggesting that this training had changed their verb biases. Ryskin, Qi, Duff, and Brown-Schmidt (2017) trained participants by showing them training trials where verb-structure pairings were accompanied by events, and found that participants used these new verb biases to interpret ambiguous sentences (see also: Ryskin, Qi, Covington, Duff, and Brown-Schmidt, 2018). Consequently, the ability to update verb biases is a lifelong process that takes into account new language experiences.

Artificial language learning experiments are another source of information about how new verbs are learned. These studies typically show that adults can learn new verb biases that qualitatively behave like a naturalistically learned verb bias. Wonnacott, Newport, and Tanenhaus (2008) trained adults using novel sentence structures and verbs, which appeared in languages that varied in whether verb-specific or language-general patterns were better predictors of structure, and whether the language generally could be characterized as having many or few alternating verbs. Wonnacott et al found that adults learned new verb biases, which they demonstrated in production, grammaticality judgments, and eye movements, and generalized what they had learned to new verbs based on the general behavior of verbs in the language. This finding also generalizes to children. Wonnacott (2011) presented five- to seven-year-old children with noun-particle word associations similar to the patterns used in Wonnacott et al. (2008). A noun-specific particle word would always occur with certain nouns, while a language-general pattern allowed nouns to occur with either particle, although one particle was more common than the other. Again, learners of the lexically-specific language learned noun biases and replicated them in generalization, while learners of the generalist language acquired the general frequency of each particle. These experiments clearly show that new statistical information can be learned in both childhood and adulthood, and that this information can be learned at multiple levels simultaneously.

Finally, humans can learn to use multiple sets of statistics about verb distributions. Twomey, Chang, and Ambridge (2016) explored sensitivity to verbs that can occur in locative structures (e.g., She filled the cup with water), but which also occur in transitive structures that are unique to that type of locative (e.g., She filled the cup vs. She poured the water). Twomey and colleagues found that adults are sensitive to these distributions, preferentially using verbs with one locative structure when they had heard the verbs previously in the corresponding transitive construction. This suggests that adults use multiple distributions of lexical cues to determine the correct verb biases for a word. Perek and Goldberg (2017) examined cases of verbs which can only appear in one type of construction, presenting participants with a language where verbs alternated between two meaningful constructions, and a similar language with one verb that appeared in only one of the constructions. Perek and Goldberg showed that adults use these new verbs based on the meaning of the construction, but were conservative with the verb that appeared in only one structure. Consequently, adults are not only able to learn from lexical distributions of verb-structure co-occurrences, but also from meaning, related distributions, and negative feedback.

The mechanism that handles verb bias learning is also sensitive to the likelihood that a particular verb will occur with a particular structure, even when that relationship has been experimentally modified. Thothathiri, Evans, and Poudel (2017) trained participants using dative structures similar to those in Coyle and Kaschak (2008), creating verbs biased toward each dative structure as well as verbs that were equally biased to each structure. Thothathiri and colleagues found that executive function played an important role in choosing sentence structures, in particular in the case of producing the dispreferred double-object dative structure with equibiased verbs. This suggests that while a verb bias is an important part of selecting a sentence structure, domain-general executive functions also play a role. Lin and Fisher (2017) also address the question of how verb bias learning works in adults, suggesting that a similar error-based learning mechanism accounts for verb bias learning in both children and adults. Consequently, the mature language production system not only retains the ability to learn new verb biases, but the learning mechanism also appears to be continuous from childhood to adulthood.

Further work suggests that learning does not only apply to statistical distributions of words and structures. Meaning at various levels – including the kind of event being described, the context in which a particular verb is used, and the semantic class to which that verb belongs – all have the potential to influence verb bias learning. First, there is evidence that meaning can condition particular syntactic structures. In an artificial language, Perek and Goldberg (2015) found that when novel structures are associated with a specific meaning, meaning competes with the observed distribution of verb-structure pairings to determine what structure participants choose to describe a scene. Thothathiri and Rattinger (2016) extended this result by manipulating whether a verb or a semantic context was a better indicator of syntactic structure, and found that whichever was the more stable cue to syntactic structure was the primary predictor of which structure an adult would likely use. Finally, this kind of learning can also be induced in a speaker's native language under the right circumstances. When a familiar English dative structure becomes more strongly indicative of a meaning than the bias of a particular verb, speakers use meaning rather than verb bias to guide their structural choices (Thothathiri and Braiuca, 2021). For instance, if the double object dative is repeatedly paired with "completed" transfer events and the majority of verbs occur in both dative alternatives, Thothathiri and Braiuca report that speakers choose to use the double object dative to describe "completed" transfer events. Jointly, these findings suggest that the kinds of meaning that sentences describe can actually be a conditioning environment for particular syntactic structures, and can outcompete lexically-dependent cues if they provide a stronger cue to structure.

Learning is also mediated by semantics at the level of individual verbs, including for different senses of particular verbs, and for verbs that belong to particular semantic classes. It is a relatively common finding that context can reverse psycholinguistic effects found in bare sentences (e.g., Nieuwland and Van Berkum, 2006). Similarly, verb biases can change depending on the context in which a sentence is presented. Hare, McRae and Elman (2003) collected verbs like "indicate" that change their biases between a direct object (indicating the door) and a sentential complement (indicating that there is a problem) based on which sense of a verb was suggested by context (see Hare, McRae, and Elman (2004) for further discussion of how verb senses interact with structural biases in corpora). When sentences were presented with context that suggested a particular verb sense, Hare and colleagues found that reading was facilitated for that sense's verb bias. A similar result was found using subjects of sentences that bias participants to expect either a transitive or intransitive sense of the same verb (Hare, Elman, Tabaczynski, and McRae, 2009). Consequently, it appears that even within a single verb, different semantic senses can create a cue to different syntactic structures. It is also the case that speakers learn about clusters of verbs with similar meanings. For example, participants rated ungrammatical sentences like "She tumbled him" as more acceptable than sentences like "She laughed him", because the semantics of "tumble" are more like verbs that can occur in a transitive sentence (Ambridge, Pine, Rowland, and Young, 2008). Ambridge and colleagues also found that participants assume that novel verbs that describe falling are constrained to the same structures as known verbs of this class. Consequently, it appears that the kinds of statistical information gathered during learning includes not only pairings of individual verbs and structures, but also diagnostic clusters of meaning such as verb senses and verb classes.

Verb biases are acquired through distributional learning, which begins in childhood and continues throughout adult life. Children can already learn about verbs before they turn two (e.g., Scott and Fisher, 2009), and show effects of learned verb biases in production by the time they are three (e.g., Peter, Chang, Pine, Blything, and Rowland, 2015; Kidd, Lieven, and Tomasello, 2006). Adults retain the ability to learn verb biases for verbs they already have biases for (e.g., Coyle and Kaschak, 2008), and in novel verb-learning tasks (e.g., Wonnacott, Newport, and Tanenhaus, 2008). These abilities are typically attributed to distributional learning, which at its most basic tracks the co-occurrence of verbs and structures (e.g., Scott and Fisher, 2009; Coyle and Kaschak, 2008). However, verb biases benefit not only from direct pairings of a verb with structures it may alternate between, but also from informative distributions of other structures (e.g., Twomey, Chang, and Ambridge, 2016), negative evidence (Perek and Goldberg, 2017), language-general statistics about the alternation behavior of verbs (Wonnacott, Newport, and Tanenhaus, 2008; Thothathiri and Rattinger, 2017), and meaning (e.g., Perek and Goldberg, 2015, Perek and Goldberg, 2017). Consequently, the effects of verb bias are modulated not only by the probability of the co-occurrence of a verb and a structure (e.g., Garnsey, Pearlmutter, Myers, and Lotocky, 1997), but also by the context of a sentence (Hare, McRae, and Elman, 2003). While verb biases are often operationalized as a verb-structure co-occurrence, their reality is somewhat more complex. These statistics are clearly created through the long-term storage of many kinds of syntactic and semantic context, which creates observable facilitation during language comprehension and production. Verb biases are both a result of language experience, and a strategy for successfully predicting the language use of others.

1.3 MODELING OF VERB BIAS USE AND ACQUISITION

Recent work has vastly expanded what is known about verb bias learning. It has shown that verb biases are not static, and that both adults and children have the capability of updating these representations (e.g., Coyle and Kaschak, 2008; Qi, Yuan, and Fisher, 2011). Humans are capable of learning distributions that predicated on multiple kinds of cues, including verbs (Coyle and Kaschak, 2008), meaning (Hare, McRae, and Elman, 2003), and verb classes (Wonnacott, Newport, and Tanenhaus, 2008). Given the complexity of verb behavior, this literature is a natural one to characterize with a cognitive model. Multiple kinds of model architectures, including Bayesian models (e.g., Perfors, Tenenbaum, and Wonnacott, 2010) and neural networks (Ambridge and Blything, 2016), have been used to characterize aspects of verb learning. These models also focus on different aspects of learning about verbs, from the initial discovery of verb classes (Perfors et al., 2010) to understanding what learned factors govern structure selection (Bresnan, Cueni, Nikitina, and Baayen, 2007). Jointly, these models suggest that learning about verbs can be characterized by relatively simple mechanisms, although there are a number of findings that are not specifically accounted for by any specific model.

One finding from the modeling literature is that models can account for the kinds of factors that may lead to the creation of verb biases. While a tendency to produce a verb in its preferred structure can be explained in terms of ease of production (Ferreira and Schotter, 2013), it is also true that a variety of semantic and discourse-level factors can be used to predict whether a speaker will choose to produce a double object or prepositional dative. For example, whether or not a particular noun is “given”, or had been mentioned previously, strongly predicts that it will be said before any non-given nouns (Bresnan, Cueni, Nikitina, and Baayen, 2007). Using features like these in combination with information about the main verb of a sentence, Bresnan and colleagues generated logistic models that they used to predict which dative a sentence would use. They found that after accounting for differences in linguistic register, this model could characterize data from both the spoken Switchboard corpus and the written Wall Street Journal corpus. Consequently, even a relatively simple model can begin to explain why some verbs tend to be associated with particular structures. More recently, this work was generalized to a much larger set of dative judgments using modern language models (Hawkins, Yamakoshi, Griffiths, and Goldberg, 2020). In this study, acceptability judgments from a larger set of verbs were predicted using models that were trained on very large corpora. Hawkins and colleagues found that these models were sensitive to the biases of dative verbs, and that they could predict dative structure about as well as the logistic models in Bresnan et al. (2007). It is important to note that these models do not use hand-selected features like the models used by Bresnan and colleagues. However, these results jointly suggest that important features for characterizing verb biases can be found using multiple different methods.

Another way of characterizing distributional learning is through the use of Bayesian models. These types of models use Bayes' theorem to update an initial probability distribution, or prior, based on new data. This updated distribution, called the posterior, can then be analyzed to understand what the model has learned. This process can occur at multiple levels – for example, a verb and a particular structure could both have their own distributions – which allows for explicit modeling of the multiple levels of learning (e.g., Perfors, Tenenbaum, and Wonnacott, 2010; Barak, Fazly, and Stevenson, 2014). Verb biases can be learned as one of these levels; however, the goal of the entire model is to understand what kinds of information need to be collected in order to successfully model human verb learning and generalization.

Bayesian models have successfully modeled a number of empirical findings, and generally suggest that verb classes in particular are a necessary component of replicating these results. A hierarchical Bayesian model with three layers can replicate the results of Wonnacott, Newport, and Tanenhaus (2008), learning either verb-specific biases or general statistics about verb classes depending on how frequent alternation is in that verb class (Perfors, Tenenbaum, and Wonnacott 2010). With the addition of the ability to cluster verbs into classes, this model also learns to differentiate alternating and non-alternating dative verbs; its ability to correctly generalize these verbs also increases as it receives more input. Finally, the addition of semantic features to the input further prevented overgeneralization. Further work has reinforced the utility of Bayesian models for understanding verb learning. While the input in Perfors et al. (2010) was limited to only dative structures, models that create verb classes can still learn the dative alternation from input that mixes dative structures and other kinds of irrelevant structures (Parisien and Stevenson, 2010). Bayesian models can also be used to model how verb generalization changes over development. Models that update verb clusters incrementally show that the ability to generalize develops over time as the model gains more knowledge about general verb classes and can move past biases in its input (Barak, Fazly, and Stevenson, 2014). Generally, Bayesian models offer insight into the levels of abstraction at which humans collect statistics, and can explicitly test whether these levels are necessary to replicate experimental findings. Importantly, they underscore the need for a level of abstraction that gathers similar verbs and learns about the behavior of the entire class.

By contrast, connectionist approaches have not produced the same kind of comprehensive models of verb learning as Bayesian approaches. However, connectionist approaches have recreated aspects of verb learning, and also have insights into the kinds of architectures that can and cannot reproduce previous experimental results. The kinds of neural nets used to create many language models use arrays of artificial neurons that collect activation, and then pass it along to other artificial neurons through adjustable weighted connections. Different arrangements of weights and neurons yield nets with different kinds of behaviors that are suited for different kinds of tasks. For example, feed-forward networks simply pass information "forward" from an input vector to an output layer, and may be used to model processes like selecting one construction from several potential options (e.g., Ambridge and Blything, 2016). By contrast, more complex architectures allow models to complete more complicated tasks, like predicting the next word in a sequence (e.g., Chang, 2002; Chang, Dell, and Bock, 2006; Elman 1990). Processes like these have produced a number of insights into how verb bias learning might proceed with less prior structure than is typically found in Bayesian models.

In principle, there is no reason why a connectionist model cannot learn to associate a verb with a particular structure. However, their architecture does not always permit the learning of verb biases even if they are able to replicate closely related effects. For example, the Dual-Path model is able to replicate many findings related to syntactic priming (Chang, Dell, and Bock, 2006). However, in part due to the fact that it separates syntax from meaning, this model does not learn syntactic information associated with particular words, and consequently cannot learn verb biases even though it implements a learning mechanism similar to the one proposed for verb bias learning. However, the Dual-Path model has been used to model the acquisition of verb classes that alternate between the English locatives (e.g., She filled the cup *with water* vs. She poured water *into the glass*) (Twomey, Chang, and Ambridge, 2014). In this study, the model was able to learn five different classes of locative verbs with varying levels of bias for each of the locative structures. A simpler connectionist model has also been shown to learn dative preferences for individual verbs in a human-like way, demonstrating early overgeneralization and naturally exhibiting behaviors like preemption (Ambridge and Blything, 2016). Preemption occurs when there are multiple possible constructions that could be used, but one is used preferentially and consequently blocks the usage of the other alternative (Goldberg, 1995; Brooks and Tomasello, 1999). One example of this is the past tense of the verb "go", which could be "goed", but is blocked by the semantically-identical form "went" (Goldberg, 1995). While this model does not perform as well as a comparable Bayesian model when presented with the same information (Barak, Goldberg, and Stevenson, 2016), it also uses a much simpler implementation than the majority of Bayesian models (Ambridge and Blything, 2016). Consequently, although connectionist models do not offer the same comprehensive solutions as Bayesian models, their simpler mechanisms offer a more parsimonious account of learning.

Generally, modeling of verb bias learning offers multiple insights into what kind of information this process potentially uses. First, allowing researchers to choose features to explain the dative alternations is about as effective as training a model on a large corpus, suggesting that both methods are sensitive to properties in the input (Bresnan, Cueni, Nikitina, and Baayen, 2007; Hawkins, Yamakoshi, Griffiths, and Goldberg, 2020). Second, Bayesian models demonstrate that learning occurs at multiple levels of abstraction, and that learning of verb classes may be particularly crucial (Perfors, Tenenbaum, and Wonnacott, 2010; Parisien and Stevenson, 2010). Finally, connectionist models suggest that learning at both the level of verb classes (Twomey, Chang, and Ambridge, 2014) and at the level of individual verb-structure co-occurrences (Ambridge and Blything, 2016) do not necessarily require complex learning processes to explain acquisition.

CHAPTER 2: COGNITIVE MODEL OF VERB BIAS LEARNING MECHANISMS

One avenue for better understanding verb biases is to carefully examine their learning mechanisms. Once basic elements of these mechanisms are understood, modeling can be used to determine specifically how they work. For verb bias learning, there are many studies that examine how different distributions affect language use. There are fewer studies that try to directly characterize verb bias learning. However, these studies have suggested that verb bias learning is both incremental and error-based.

Incremental learning refers to learning that proceeds one trial at a time. Unlike learning that benefits from observation or insight, incremental learning gradually strengthens the association between a stimulus and an outcome. Crucially, this also means that this kind of learning is undone in a trial-by-trial manner as well. This insight is the core of a paradigm called reversal learning. In reversal learning, participants are trained on a rule, often to a specific level of performance, or criterion, and then are trained on the opposite of that rule (Izquierdo, Brigman, Radke, Rudebeck, and Holmes, 2017). Rules vary depending on the species and task but could include discriminating between two different shades of gray (Hoffmann, Perkins, and Calvin, 1956), learning to turn right or left in a maze (McDaniel, 1969), or even learning categories of subtly different Gabor patches (Cantwell, Crossley, and Ashby, 2015). Recovery from the reversal may have one of two outcomes. In fast reversal, learning from the initial rule allows participants to learn what dimensions of a particular problem make up the rule, and when the rule is reversed, participants quickly learn to use those same dimensions to make an opposite response (Sanders, 1971; see Kruschke, 1996, for review). For instance, human participants in the gray-discrimination task might realize that they are rewarded for choosing a particular shade of gray, and simply switch to choosing the other shade when the rewards change, causing them to reach a criterion much more quickly in the second block. Another possible outcome is a slow reversal. Under these conditions, previous learning does not improve learning the reversed rule; instead, it is common for participants only gradually begin to give the opposite response (e.g., Hoffmann, Perkins, and Calvin, 1956). The rats in Hoffmann, Perkins, and Calvin consistently took more than twice as many trials to learn a reversed gray discrimination, suggesting that their earlier learning may have actually impeded reversal. The "fast" and "slow" in these situations refers to how quickly the criterion is reached – in the first case, more quickly than the first time the rule was learned, and in the second, more slowly.

Converging evidence from neuroscience and behavioral studies can help further characterize learning by explaining developmental and evolutionary gradients in reversal behavior. Behavioral work shows two overall trends. First, more children reverse slowly than adults (e.g., Kendler and Kendler, 1970). Second, slow reversal is more common among animals of lower order taxa, such that rats are more likely to reverse slowly than humans (Sanders, 1971). These findings are complemented by findings from neuroscience. Reversal learning is impaired in patients with ventromedial prefrontal cortex damage (Fellows and Farah, 2003). These areas support the representations that allow learners to compare trials and flexibly shift strategies based on aspects of those sets of trials (Izquierdo, Brigman, Radke, Rudebeck, and Holmes, 2017; Murray and Gaffan, 2006). If prefrontal areas are impaired, then reversal learning is handled by subcortical structures like the basal ganglia, which requires multiple trials to learn to inhibit previously learned, habitual responses (Frank and Claus, 2006). Consequently, fast reversal occurs when a learner is able to represent a rule and compare across trials (Izquierdo et al., 2017); this is easier for adults than children, and for humans than for animals. Slow reversal occurs when trials cannot be easily represented and compared, and learning is handled primarily by subcortical structures that require multiple trials to re-learn reward structures (Frank and Claus, 2006). These results also allow us to better understand why there are two different kinds of reversal learning. Fast learning is characteristic of the cognitive flexibility derived from the ability to easily compare trials, while slow learning is the result of incrementally unlearning habits on a trial-by-trial basis.

Linking verb bias learning to other reversal learning studies helps determine whether verb bias learning is incremental or not. Kelley (2019) found incremental verb bias learning in two different tasks. In both studies, participants learned new biases for six verbs, three transitive and three dative. In each of these groups, two verbs were biased toward new structures, while one verb appeared an equal number of times in each of its syntactic alternates. For instance, the dative verb "give" would be biased toward the double object dative structure, the verb "hand" would be biased toward the prepositional dative, and the verb "show" would appear in both dative structures an equal number of times. Transitive verbs were theme-experiencer verbs, which are more likely to alternate between active and passive structures (Ferreira, 1994). The dative verbs selected also alternate regularly between the double object dative and prepositional dative. Both of these studies also shared a block structure, which would reverse the associations between verbs and structures in the second block. If a participant learned to produce "give" in the double object dative in the first block, they would then learn to produce "give" in the prepositional dative in the second block. The two experiments differed in the procedures used to induce learning. One adapted a method used in Potter and Lombardi (1998), presenting sentences using rapid serial visual presentation, and inducing production by asking participants to repeat the sentences after a short interval. The other adapted the paradigm used in Coyle and Kaschak (2008), training participants by asking them to complete sentence stems with pre-determined structures, and testing with separate trials where participants were allowed to produce sentences with any structures they wanted. However, both experiments found that learning was less successful in the second block than the first block, and that this was true only for dative verbs. In other words, this paradigm showed slow reversal, where performance at learning the reversal is less successful than initial learning. This was attributed to an inability to use attention to detect relevant dimensions of the stimuli, either due to the complexity of tracking the changing verb biases across multiple trials, or because language production encourages procedural learning. Consequently, verb bias learning should be modeled using an incremental mechanism.

Similarly, saying that verb bias learning is error-based makes specific predictions about how learning should behave. Error-based learning refers to circumstances where learning is greater when there is more error, such as would occur when the system encounters a more unexpected event. When the event is more likely, the system learns very little. However, when the event is uncommon, the system adjusts its parameters significantly in order to accommodate predicting that event more frequently. In language production, error can potentially arise from many sources, but one of the most relevant for verb bias learning is frequency. Syntactic priming, or the likelihood that a structure will be used again by a speaker, is greater when a structure is less common (Bernolet and Hartsuiker, 2010). Bernolet and Hartsuiker also found that priming is strongest when a prime sentence uses a less common structure and a verb that is biased against that structure. In other words, less-predictable structures and verb-structure pairings resulted more learning, which is compatible with an error-based learning account of syntactic priming. A cognitive model should be able to replicate this kind of behavior.

Showing that verb bias learning is error-based involves measuring differences in learning that depend on that verb's preexisting bias. Lin and Fisher (2017) tested this by presenting participants with dative verbs that were either more biased toward the prepositional dative or the double object dative and training these verbs with or against that bias. For a verb like “give”, which is more likely to occur in the double object dative, with-bias training would involve presenting “give” in double object dative sentences, while against-bias training would use prepositional dative sentences. Lin and Fisher found that this learning was error-based at two different levels. First, more verb bias learning occurs when participants are presented with verbs in structures that are not their preferred structure. Additionally, this effect is modulated by the frequency of a syntactic structure. For example, more learning occurs when a verb is biased toward the less-common double object dative than when it is biased toward the more-common prepositional dative. These two effects interact, such that learning about a dispreferred DO structure causes the greatest amount of learning, while biasing a verb toward a preferred PD structure creates very little learning. Consequently, these results suggest that

The following simple model can account for both the error-based (surprisal) and the reversal-learning effects using a learning mechanism that is both incremental and error-based. This model contains three feed-forward layers. As can be seen in Fig. 1, at the input layer, the model receives a one-hot vector that codes for a single verb. This information is then passed to a three-node hidden layer. Finally, this information is passed to a two-node output layer, where each node represents a syntactic structure. Using this architecture, the model was asked to learn associations between input verbs and output structures. One set of data mirrors the stimuli used in Experiments 1 and 2 from Kelley (2019), asking the model to first learn an association between a particular verb and structure, and then to reverse that association. For example, “hand” might first appear only in the prepositional dative, and then only in the double object dative. The second set of data mirrors the procedure used in Lin and Fisher (2017), which selected verbs with different biases toward dative structures. To mirror this, the different verbs were first given different baseline "biases" – for example, "throw" might only appear in the prepositional dative, while "give" would appear in both datives half the time. Then, the verbs were trained toward either the structure they preferred – for example, "throw" would be trained further toward the prepositional dative – or toward the structure they dispreferred.

2.1 STUDY 1: REVERSAL LEARNING

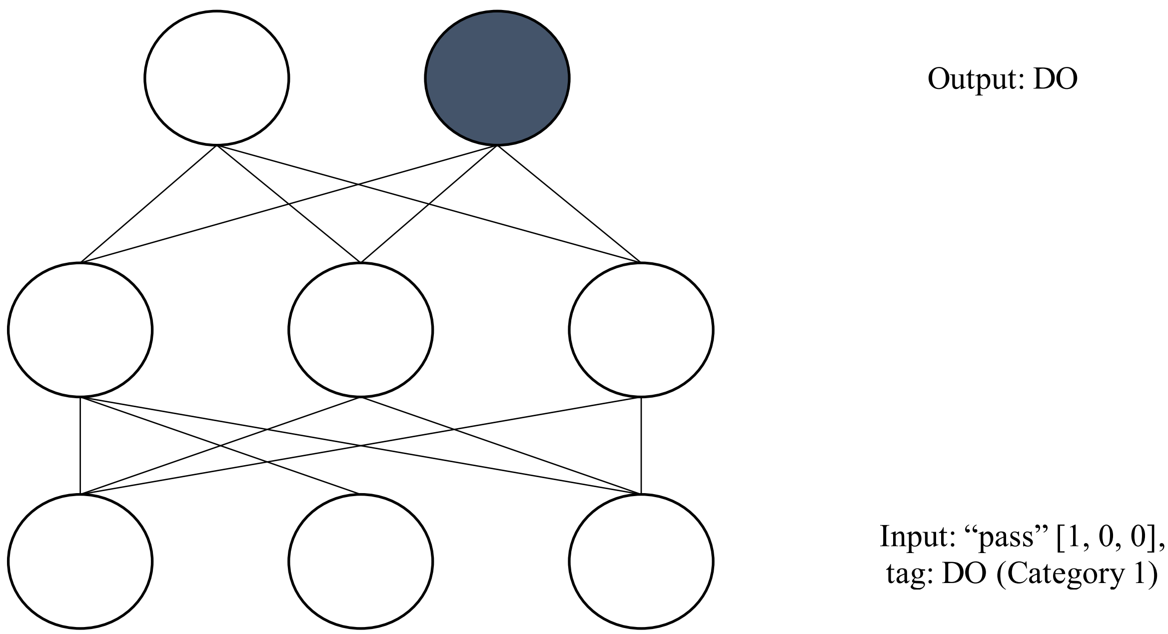

Figure 1: A schematic of model with sample input and output.

2.1.1 Methods

A three-layer feed-forward model was implemented using PyTorch (Paszke et al., 2019). Learning was done via backpropagation, and loss was calculated using a cross-entropy loss function, which allows for discrimination between two or more categories (Torch Contributors, 2019). The learning rate of this model was 0.05, and it did not use momentum. The hidden units of this model used a hyperbolic tangent activation function, while activation at the output units was evaluated using a softmax function.

Input to this model was individual verbs, represented by simple one-hot vectors. An example of this type of representation, for the verb “pass”, can be seen in in Fig. 1. Associations between these vectors and particular syntactic structures by designating each structure as a category; for instance, the prepositional dative was Category 0, while the double object dative was Category 1. Although this model only required two categories, further structures could be added by increasing the number of categories. For instance, a model with three structures would contain Categories 0, 1, and 2. Consequently, the model learns to categorize each vector, treating verb biases as a kind of probabilistic category.

The model was trained by first setting all weights to 0.5, in order to mimic training a verb that is not biased toward any particular structure. Verb vectors were then presented to the model in a random order, and the model was updated stochastically after each presentation. In order to simulate reversal, the model was first presented with two verbs, one of which only appeared in the prepositional dative, and one of which appeared only in the double object dative. As in Experiment 2 of Kelley (2019), the model saw 10 presentations of each verb in the first block, for a total of 20 trials. After this, the contingencies learned in the first block were reversed. For example, if “hand” was presented in the prepositional dative in the first block, it was presented in the double object dative in the second block. These reversed contingencies were presented to the model exactly as they were in the first block. The process of setting the weights to 0.5, presenting the first block, and presenting the second block was simulated 100 times, to represent individual participants in a study.

2.1.2 Results

In order to evaluate learning in the model, the bias toward a particular structure was measured each time the model saw a verb and updated its categories. For example, if the model saw "hand", once backpropagation was complete, the model's structural biases for both "hand" and "send" would be evaluated. The bias of these verbs was measured by applying a softmax function to the linear output units of the model. This process allows the activation of the units to be translated into percentages that indicate how likely it is that the particular input will be assigned to that category (Duda, Hart, and Stork, 2001). For example, if "hand" results in an output of 0.7 for Category 0, that is a likelihood of 70% that "hand" will be assigned to the prepositional dative. Biases were recorded only for the correct verb-structure pairings.

The average of these biases for each block is presented in Fig. 2. In Block 1, the average bias toward the correct structure was 0.606, while in Block 2, the average bias toward the correct structure was 0.456. Evaluated using a paired t-test, performance in Block 1 was significantly better than in Block 2 ($t=6.66$, $p<0.05$). These results conform with the findings in Kelley (2019), and show that learning in Block 2 likely includes unlearning the biases acquired in Block 1. In other words, it mimics human behavior, showing that learning in the second block is impaired by learning in the first block.

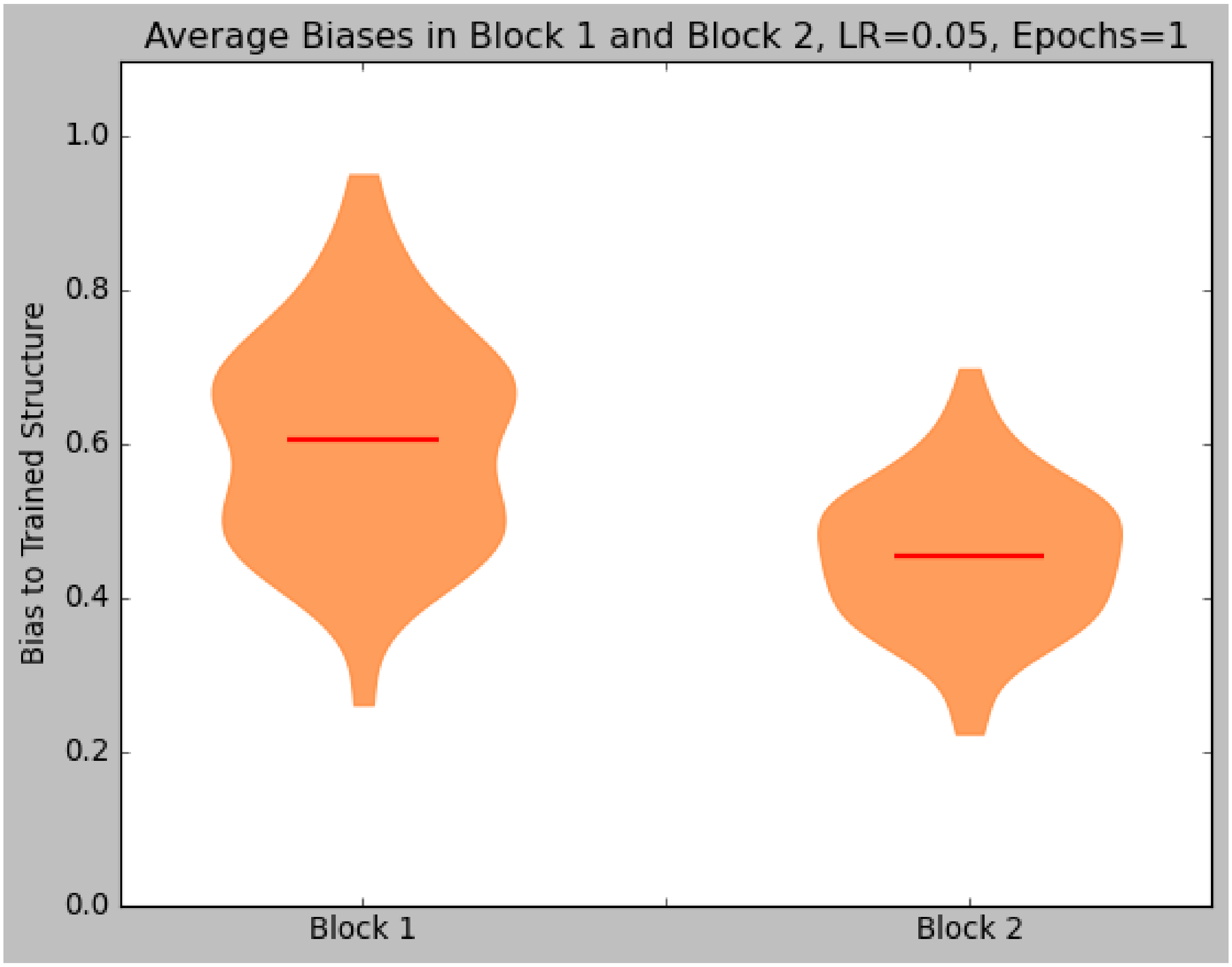

Figure 2: Comparison between average bias toward trained structure in Block 1 and Block 2.

2.2 STUDY 2: SURPRISAL

2.2.1 Methods

The model used to evaluate surprisal is the same as the model used to evaluate reversal learning. Input to the model was the same one-hot vectors used in the reversal learning simulation. Training the model to demonstrate surprisal, however, had two distinct phases: an initial "life experience" phase, and then a training phase where the model was trained either with or against the bias of a particular verb.

In the life experience phase, the model's weights were initially set to random weights. The model was then presented with two hundred epochs of experience with three "verbs", which are labeled with English verbs for the sake of readability. The percentages in with which each verb appeared in each structure can be seen in Table 1. Although the distributions for these three verbs are arbitrary, they do generally conform to distributions seen in English. One verb, "pass", only appears in the prepositional dative. The second verb, "hand", appears in the prepositional dative 70% of the time, and in the double object dative 30% of the time. Finally, "give" appears in both structures 50% of the time. Consequently, this creates a life experience where each verb has an individual bias, and the prepositional dative is overall more common than the double object dative.

A training phase is conducted after a life experience phase. In the training phase, the verbs "pass" and "hand" were biased in the same way as they were for the participants in Lin and Fisher (2017). Half of the training phases were with-bias, and half were against-bias. In a with-bias training condition, verbs were further trained toward the biases they already had, so that "hand" occurred 10 times in the prepositional dative, and "give" occurred 10 times in the double object dative. In the against-bias condition, the training was always counter to the verb's current bias, so that "hand" received 10 trials of double object dative training and "give" received 10 trials of prepositional dative training. After each training trial, the bias of the model toward each structure was evaluated, and the average probability of producing a double object dative was obtained using a softmax function (Duda, Hard, and Stork, 2001).

An individual simulated participant would first receive life experience, then against-bias training, then receive life experience again, and finally receive with-bias training. Forty total participants were simulated.

Table 1: Distributions of Structures for Each Verb in Surprisal Simulation

Verb	Percentage Prepositional Dative Structures	Percentage Double Object Dative Structures
Pass	100%	0%
Hand	70%	30%
Give	50%	50%

2.2.2 Results

Lin and Fisher (2017) report the following effects: a training effect, a verb bias effect, and a surprisal effect. The training effect means that both adults and children were more likely to produce a structure they were trained to produce; for example, being trained to produce double-object datives meant that participants were more likely to produce double-object datives in that condition. Both adults and children also showed evidence of a verb bias effect, demonstrating a greater likelihood of producing the structure that a verb was originally biased toward with that verb. Finally, they demonstrated surprisal by comparing PD training to DO training. Lin and Fisher found that with-bias PD training was significantly different from against-bias PD training, but that with-bias and against-bias DO training were not significantly different from each other. From this, they concluded that surprisal worked at two different levels: the level of the individual verb, and at the level of syntactic structures.

In order to compare the behavior of the model with the effects found in Lin and Fisher (2017), the between-subjects t-tests that compared the two PD-trained verbs and the two DO-trained verbs were replicated. Like Lin and Fisher, the two PD-trained verbs had significantly different DO biases from each other (Hand: 0.28, give: 0.45, $t(39)=-12.82$, $p<0.05$). Unlike the results from Lin and Fisher, the two DO-trained verbs also had significantly different DO biases (Give: 0.58, hand, 0.45, $t(39)=-10.11$, $p<0.05$). The model results are presented in Figure 3.

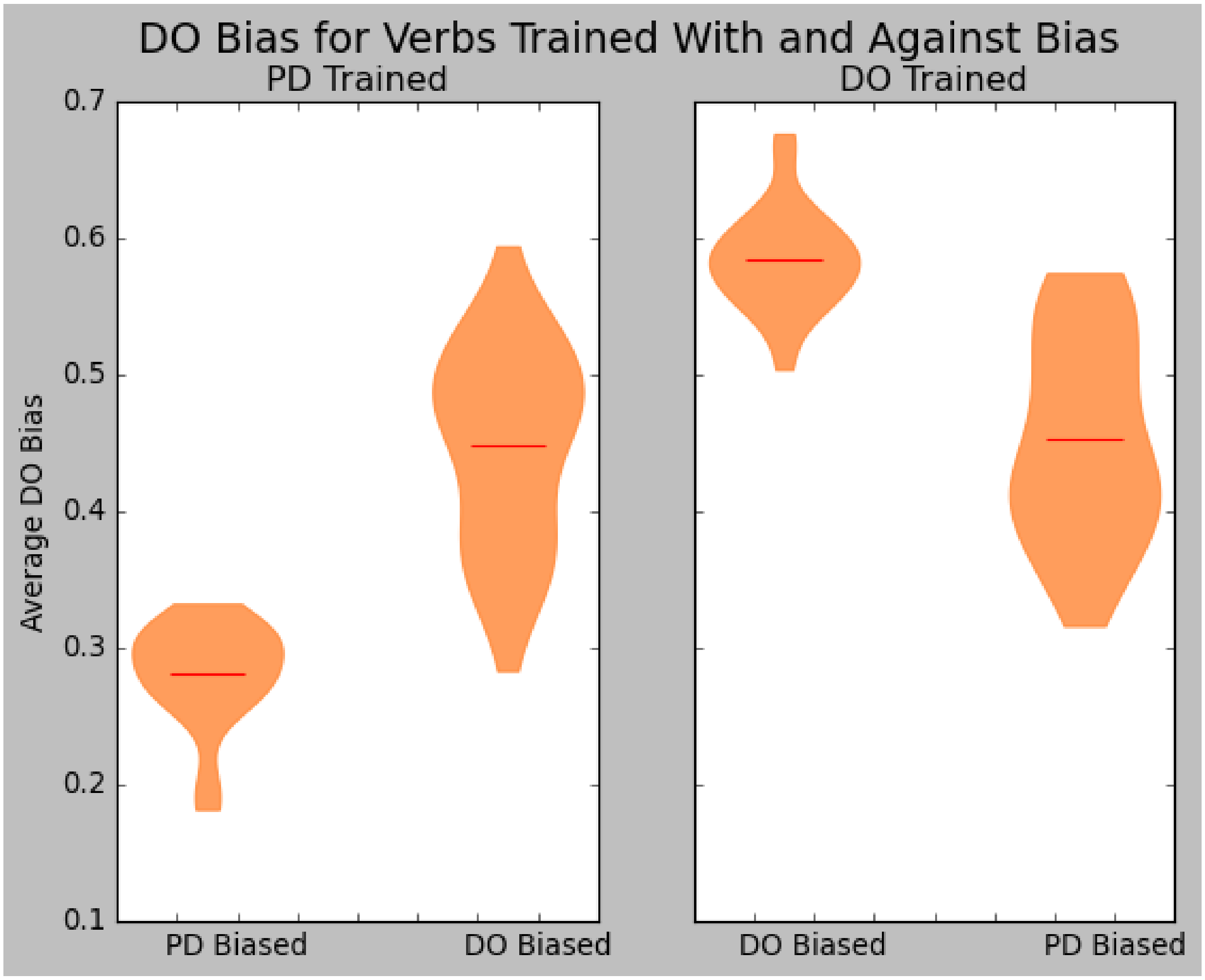

Figure 3: Model results replicating Lin and Fisher (2017).

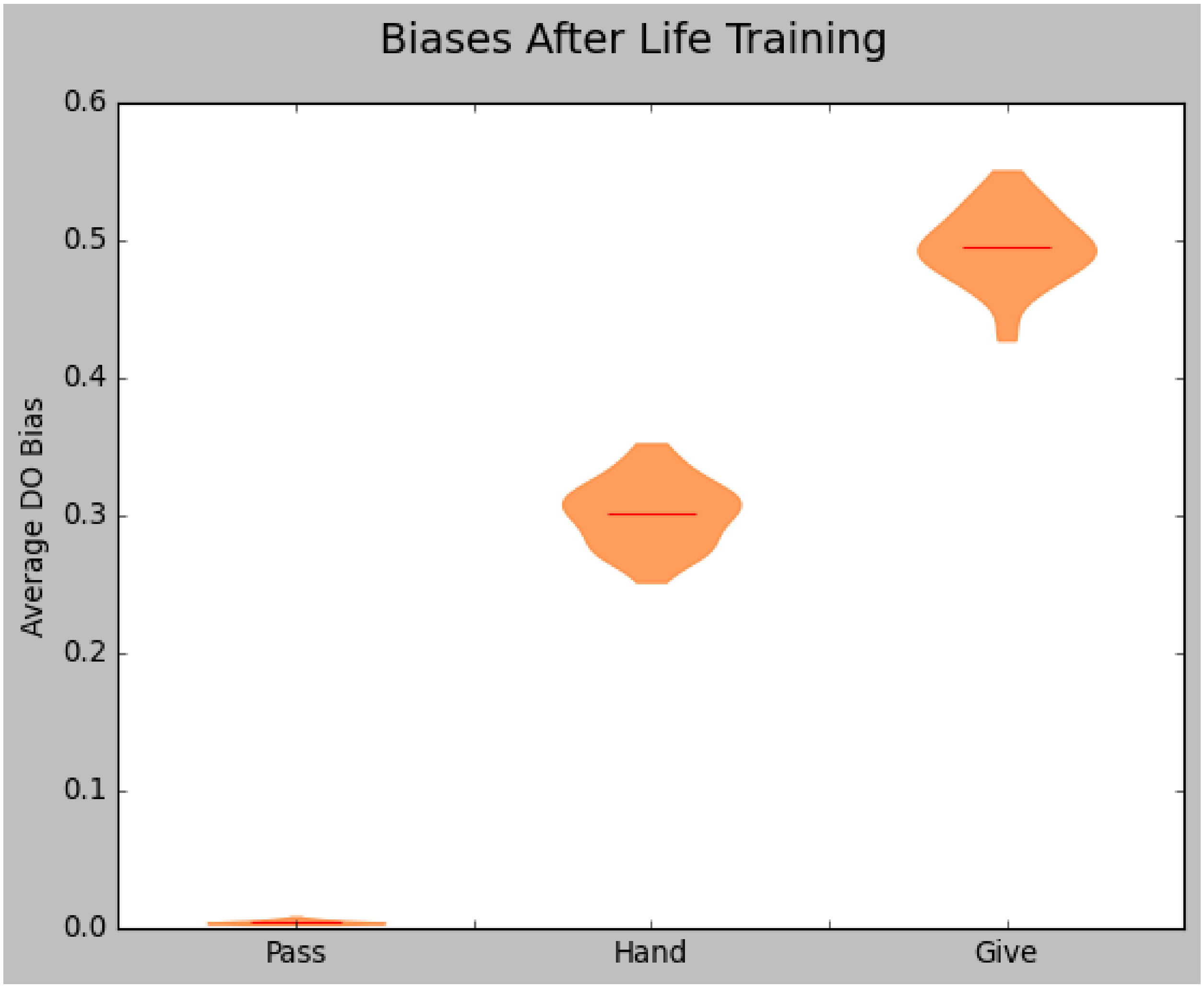

Figure 4: Model biases after life training but before experiment training.

Although these results are technically different than those from Lin and Fisher (2017), it seems that the pattern of significance does not fully capture the similarities between the model and the behavioral data. First, when the effect size is calculated for the PD-trained and DO-trained verbs, the model's effect size is larger for the PD-trained verbs (d=2.97) than for the DO-trained verbs (d=2.25). Additionally, both are large effects. One possible conclusion that can be drawn from this is that while the model only relies on distributional learning to choose structures, speakers are also affected by many other aspects of the sentences they produce (e.g., Bresnan, Cueni, Nikitina, and Baayen, 2007). From this perspective, one interpretation is that the model is simply less noisy than the human subjects, but that both produce a similar numerical pattern of results. This can be confirmed by comparing the model graph above to the Figure 5 below, which approximately reproduces data from adult participants in Lin and Fisher. Visually, the pattern of results found by the model and the results reported in Lin and Fisher are quite similar.

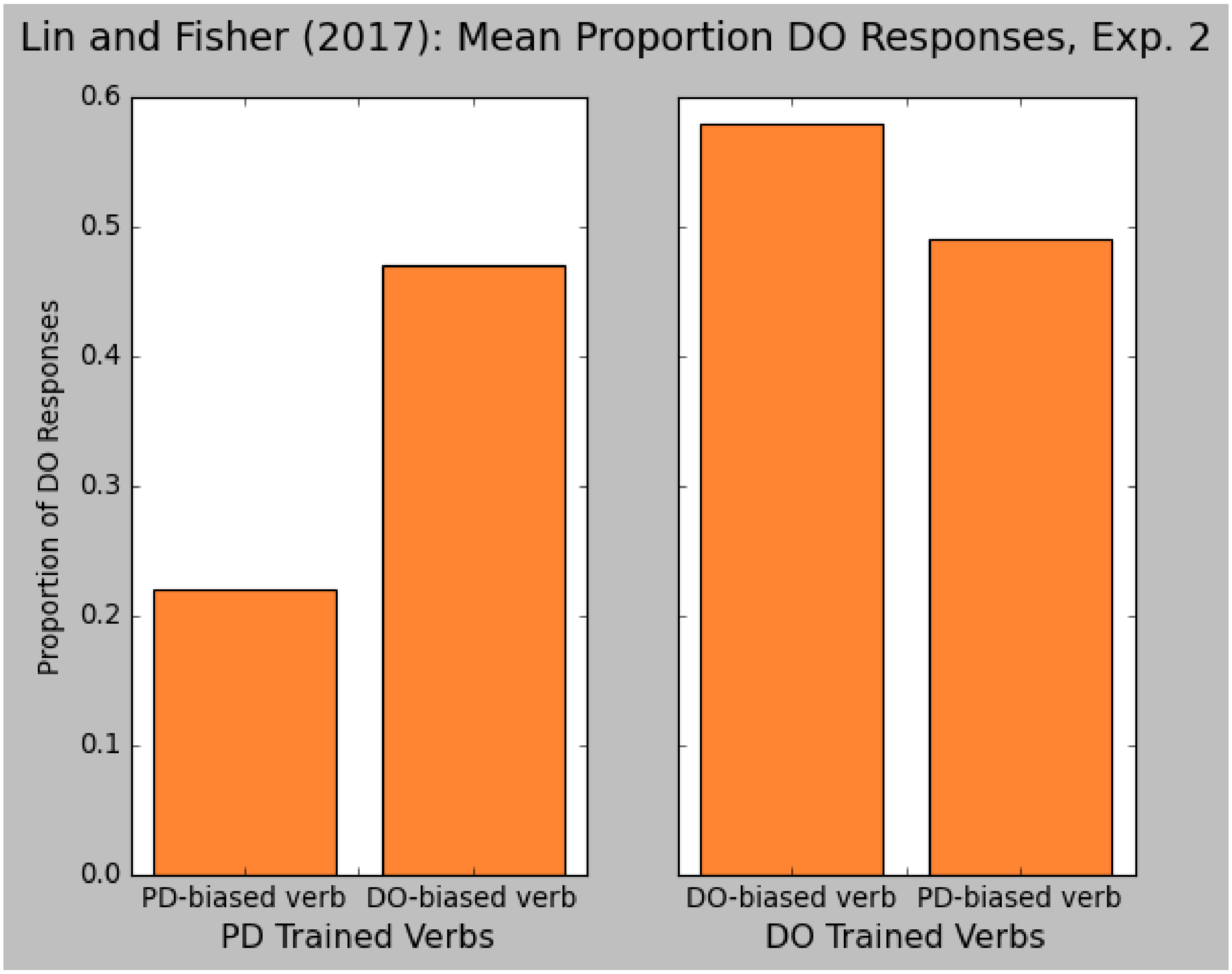

Figure 5: Estimated means from adult participants in Lin and Fisher (2017), Study 2.

Finally, does the model support the claim that this pattern of data can only be obtained by surprisal acting at both the level of individual verbs and the level of structural frequency? This point is actually somewhat ambiguous in Lin and Fisher (2017). Although “hand” and “show” are more DO-biased than other dative verbs, in absolute terms they are approximately equally likely to appear in both structures, or even slightly biased toward the prepositional dative (see

Lin (2020) for further details on norming structural preferences for these verbs). Because the verbs in this study are not truly biased toward the double object dative, their individual biases could account for the finding of greater learning in the double object dative. In other words, because the slight PD bias of the verbs themselves is confounded with the overall PD bias of English, it is possible that the findings of Lin and Fisher can be explained only by surprisal at the level of the verb, rather than at the level of the verb and structure. This limitation exists in part because of the limited set of English dative verbs. Although it can be difficult to consistently norm dative verb biases, common alternating dative verbs generally do not occur primarily in the double object dative (Lin, 2020). Consequently, it is difficult to select alternating verbs of English that represent a full range of biases for each structure.

However, since the model can replicate the findings from Lin and Fisher (2017), it can potentially begin to show whether surprisal is necessary for these effects. Generally, the model results seem to indicate that it is not. Early in training, the model exhibits a general preference for the prepositional dative. Figure 6 shows the estimated bias for a novel verb after one epoch of the life experience phase described in the methods section. This simulation was carried out 50 times. All other aspects of the model are the same, except that it has one additional input node to allow for the presentation of a novel verb. The model clearly shows an overall bias away from the double object dative and toward the prepositional dative.

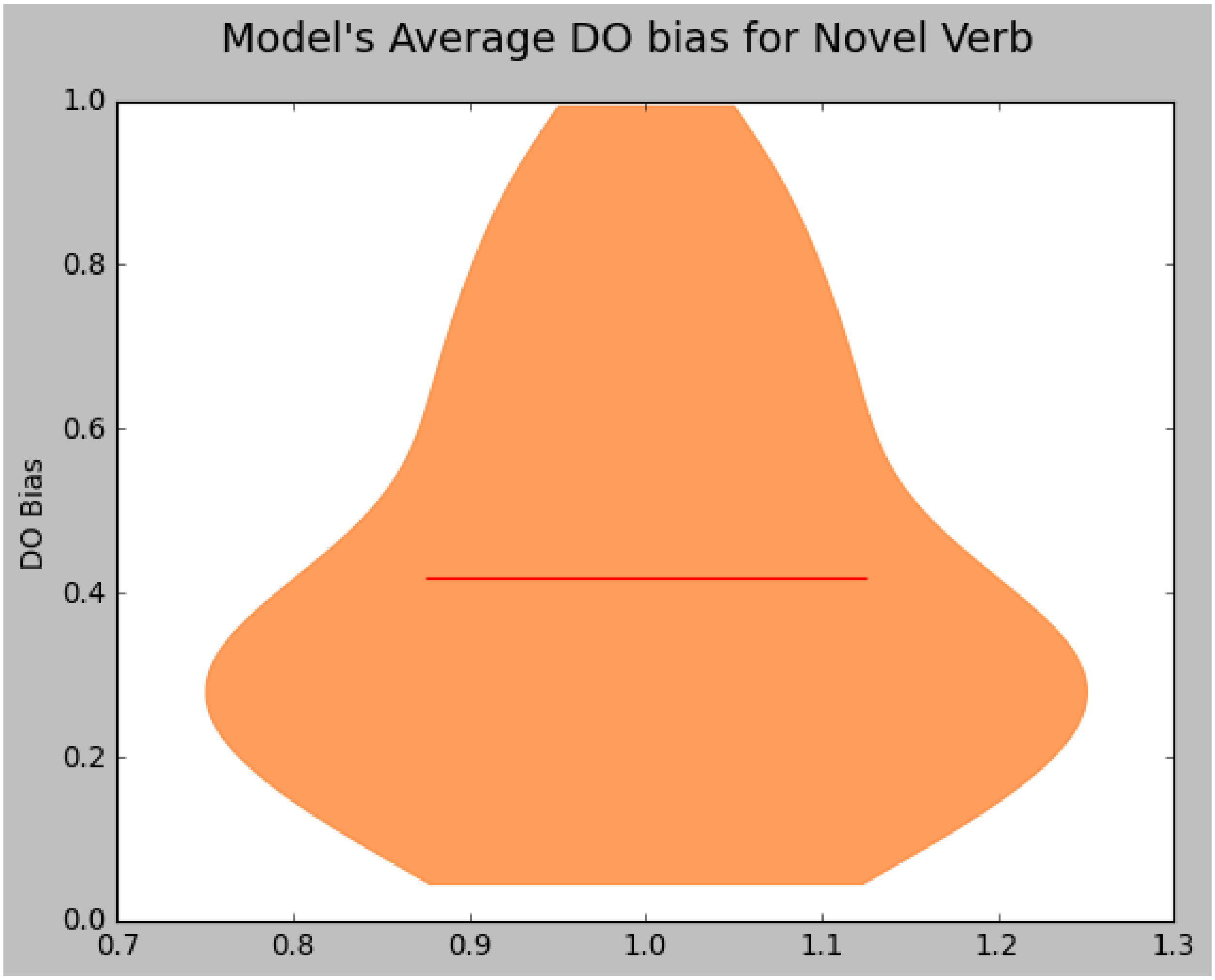

Figure 6: Model's average DO bias after one epoch of life experience.

However, the model eventually learns that verbs are better predictors of biases and suppresses the overall distributional tendency toward the prepositional dative. Below in Fig. 7 are the results of a version of Lin and Fisher (2017) that uses two verbs with symmetrical biases. One is a verb that maintains its 70% PD bias; the other is a verb with a 70% DO bias. In order to maintain the same overall bias toward the PD as in the original study, a fourth verb was added, and the model had one additional input node. Values for these proportions can be seen in Table 2. Otherwise, all methods are the same as those described in the methods section above. The magnitude of the difference in average DO bias between the PD-trained and DO-trained conditions are nearly identical. For the PD-trained condition, the difference is 0.32 (PD-biased mean: 0.25, DO-biased mean: 0.57). For the DO-trained condition, the difference is 0.33 (PD-biased mean: 0.42, DO-biased mean: 0.75). If surprisal depended on structure and verb bias, then the DO-trained PD-biased verb should have exhibited greater surprisal than any other condition, and changed its bias considerably. However, we instead see both conditions moving slightly when they are trained with their bias, and dramatically against their bias. Moreover, the similar magnitude of the change suggests that these results support surprisal at the level of an individual verb but not at the level of structures.

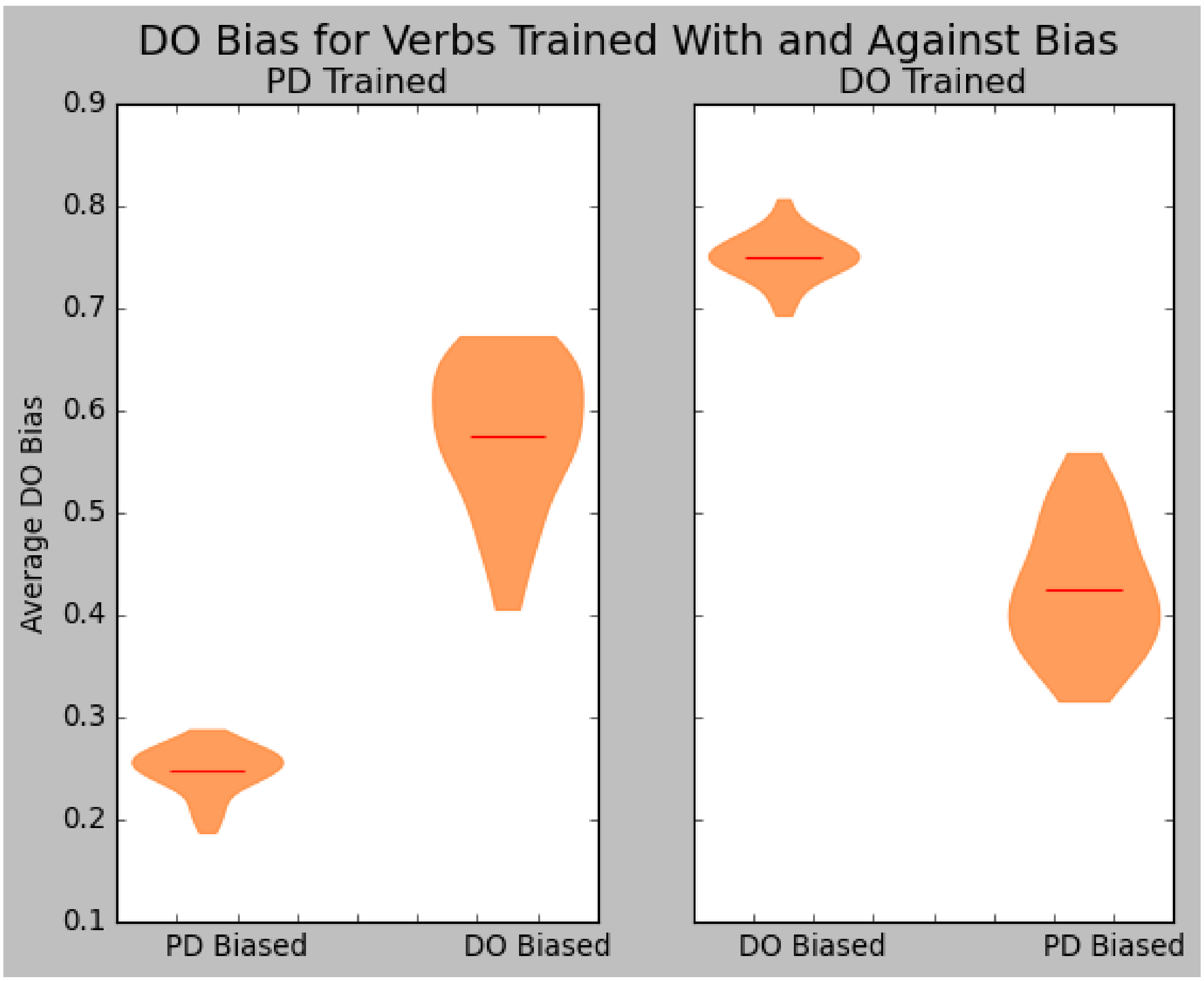

Figure 7: Model results using symmetric verb biases in Lin and Fisher (2017) paradigm.

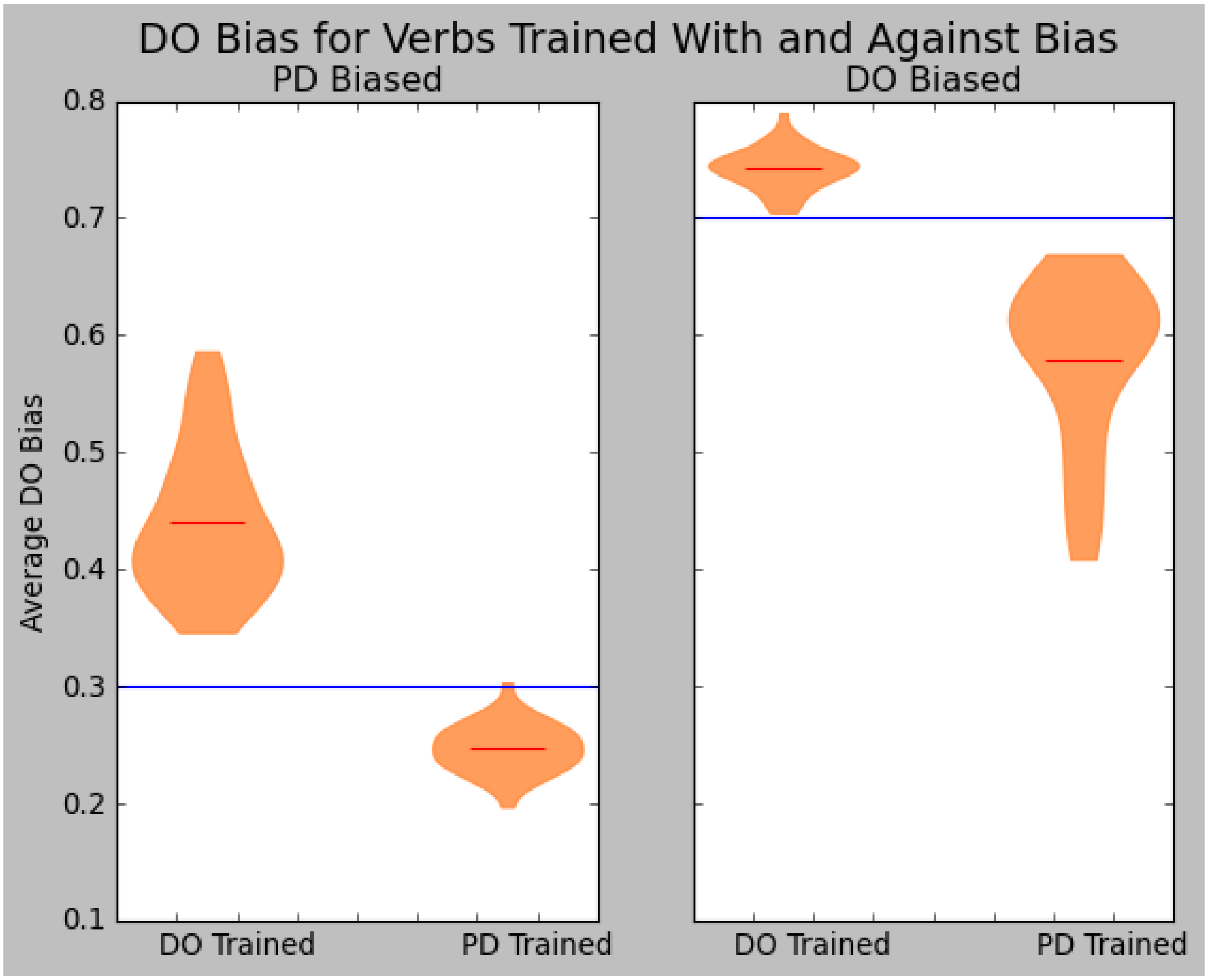

Figure 8: Re-graphed results from Fig. 7, demonstrating that the against-bias training surprisal is the same distance from baseline for both verbs.

Table 2: Distributions of Structures for Each Verb in Surprisal Simulation

Verb	Percentage Prepositional Dative Structures	Percentage Double Object Dative Structures
Pass	100%	0%
Throw	90%	10%
Hand	70%	30%
Give	30%	70%

The final investigation of surprisal examined how the model addressed how the model responded to novel verbs. In this case, the model received life training for three verbs, identical to the biases seen in Table 1. However, the model was then experimentally trained on two verbs that it had never seen before. This particular version of the model was instantiated with five input units, so that there was no overlap between the features of the three life experience verbs and the two novel verbs. During experimental training, one novel verb ("moop") was trained only in the DO, while the other novel verb ("blick") was trained only in the PD. As can be seen in Figure 9 below, the model does show structural surprisal for the DO-trained novel verb when compared to the PD-trained novel verb (Average change in DO bias from baseline for DO training: 0.21, Average change for PD training: -0.02, $t(39)=11.50$, $p<0.05$). Consequently, the lack of surprisal seen in trained verbs is not the result of a lack of learning about structural frequency. Instead, it is evidence that the model relies on the highly-predictive biases of individual verbs when they are available, and uses structural frequencies only when biases are not available.

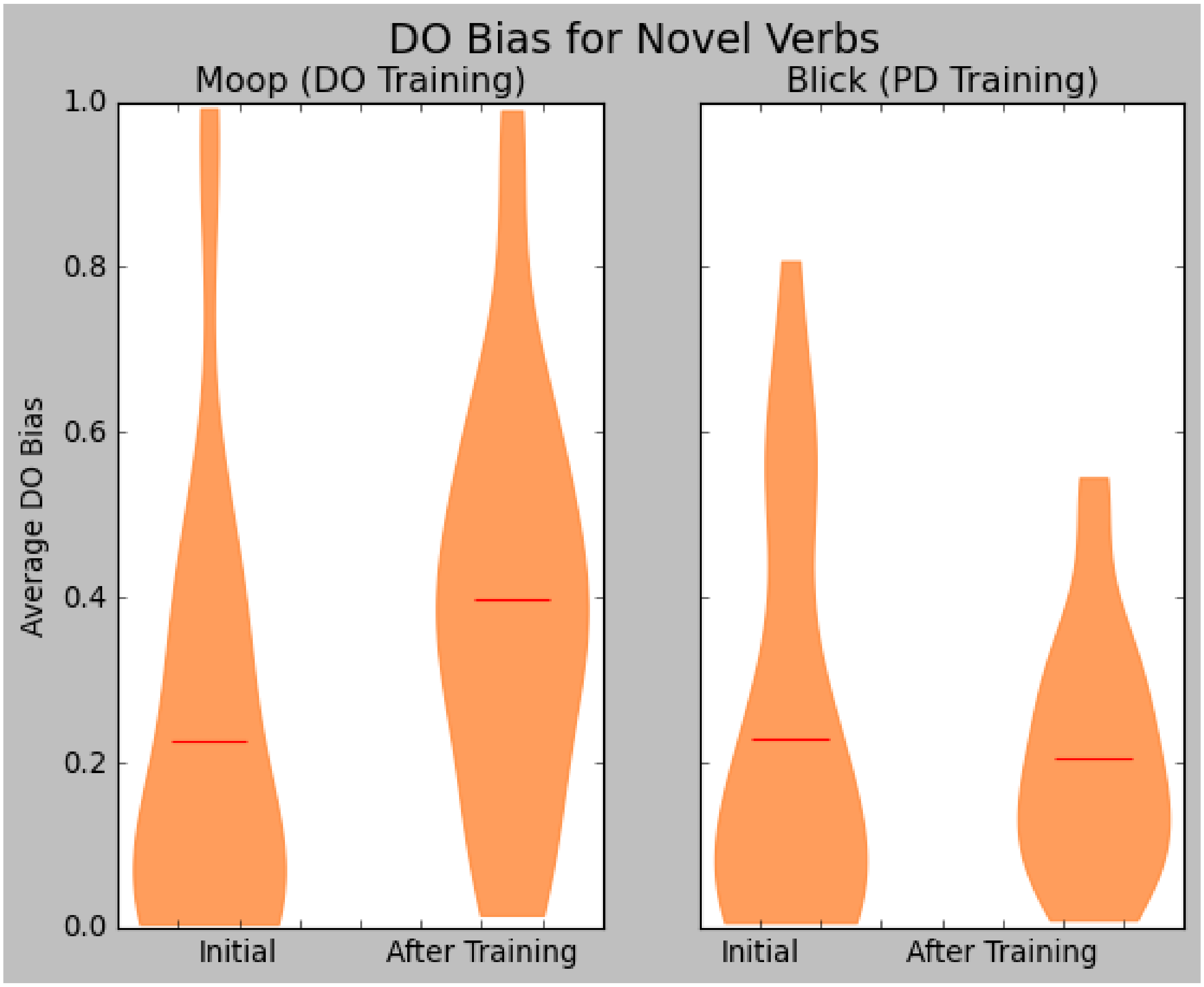

Figure 9: Model biases after experimental training for two verbs that never received life experience, demonstrating surprisal for DO but not PD training.

2.3 GENERAL CONCLUSIONS

Overall, the model successfully replicated the effects reported in Kelley (2019) and in Lin and Fisher (2017). Specifically, this model demonstrates incremental slow reversal as well as surprisal that is graded by verb bias. In other words, this model is sensitive to the same types of statistics as humans are and learns about them in a similar manner. However, this model does not fully support the conclusion that the results presented in Lin and Fisher (2017) require sensitivity to the overall frequency of syntactic structures. Rather, it appears that the model can replicate these results without sensitivity to overall structural frequency for verbs that it already knows.

This does not necessarily mean that the model has failed to replicate human behavior. Because the DO-biased verbs in Lin and Fisher (2017) are less DO-biased than the PD-biased verbs are PD-biased, it is ambiguous whether their results reflect greater surprisal for both unexpected verbs and unexpected structures. This is also true for studies that find greater syntactic priming in the prepositional dative. For example, Bernolet and Hartsuiker (2010) report greater overall DO priming, as well as surprisal based on the biases of individual verbs. However, for the group of Dutch verbs that they used, the majority of them are also PD-biased. Consequently, it is again unclear whether the effects they report are separate, or due to an overall PD bias in their stimuli. Additionally, humans are known to weight verb bias cues more or less strongly depending on how predictive they are. For example, Thothathiri and Braiuca (2021) demonstrated that with enough training, participants would begin to use sentence meaning rather than the bias of individual dative verbs to choose a structure. One possible interpretation of the model is that it is acting like a human that is very strongly reliant on verb-specific cues. Because this weighting is based on input, even if it turns out that humans do demonstrate surprisal at both the levels of verbs and structures, it is possible that the model could also learn to appropriately weight these cues with more naturalistic input.

Additionally, similar models have been used to explain results in the verb learning literature. Ambridge and Blything (2015) used a similar connectionist feed-forward model to demonstrate a number of effects that had previously been observed in the verb learning literature. Ambridge and Blything allowed their verbs to participate in three structures rather than two, and their verb vectors were derived from semantic ratings of each verb. However, like the model described above, their model contained only one hidden layer, and received the task of predicting a structure given a verb representation. Using this slightly different architecture, they found that teaching their model which verbs appeared in specific structures followed a similar trajectory to children, and that the fully-trained model gave verb judgments that were similar to those of adults. Finally, this model also seems to generalize to novel verbs like human participants. This includes demonstrating learning patterns like preemption, in which the specific use of a structure with a verb that could alternate blocks use of the potential alternate structure (e.g., Goldberg, 1995?). Simple connectionist models can explain a variety of verb bias results, suggesting that they may provide a way forward for explaining other findings in this literature.

Taken together, the results reported in this chapter and the results from Ambridge and Blything (2015) suggest that the architecture presented here may be sufficient to explain many results in the verb bias learning literature. These models do rely on having a definitively-labeled structure for each sentence, and assume representations for individual verbs and structures. However, the model does suggest that the results from Kelley (2019) and Lin and Fisher (2017) share a single, simple learning process. Additionally, while the Blything and Ambridge results rely on semantic representations of these verbs, the results replicated in this chapter do not require semantic representations. While a more complex architecture may be needed to explain other effects in this literature, reversal learning and surprisal seem to only require some representation of a verb and some feedback about which structure it has occurred in. Otherwise, these effects arise naturally from the structure of the input data and properties of the model.

CHAPTER 3: TRANSFER BETWEEN SIMILAR VERBS

Statistics are collected and updated at multiple levels during verb bias learning. Studies of verb bias learning in adults typically focus on changing the association strength between a single verb and structure. However, it is clear that speakers track more than this single co-occurrence statistic. Studies show that verb biases interact with verb sense, suggesting that different senses of a verb may be able to maintain separate biases (Hare, McRae, and Elman, 2003). Another set of studies investigates how children and adults balance semantic and statistical information about verbs, which affects how individual verbs are generalized (e.g., Ambridge et al., 2008), and how entire verb classes behave (e.g., Wonnacott, Tanenhaus, and Newport, 2008). Ultimately, these literatures suggest that verb biases are one of many attributes that affect the behavior of verbs, and that all of these components are necessary to explain lifelong language learning.

The primary process investigated by adult studies of verb bias learning is the strengthening of the association between a specific verb and a specific structure. These findings are important for understanding when and how verb biases are learned. Adults can change biases for verbs they already know through exposure to skewed distributions in both production (Coyle and Kaschak, 2008) and comprehension (Ryskin, Qi, Duff, and Brown-Schmidt, 2017). Children can also adjust their biases for known verbs (Qi, Yuan, and Fisher, 2011). Verb bias learning appears to be error-based, in that exposure to a verb in its dispreferred structure causes more learning than exposure to a verb in the structure that it already prefers (Lin and Fisher, 2017). Error-based learning is a lifelong process that is used to gradually adjust how strongly verbs are associated with specific structures (see e.g., Chang, Dell, and Bock, 2006; Dell, Kelley, Hwang, and Bian, 2021, for models of error-based language learning).

As mentioned previously, a simple association between a single verb and a single structure cannot fully explain how verb biases influence behavior. The first additional factor is how verb biases interact with verb sense. Different senses of a verb can have very different subcategorization biases; for example, for the verb *indicate*, the difference between an alarm *indicating* a problem and a person *indicating* the door are two discrete senses of this verb (Hare, McRae, and Elman 2004). When preceding context suggests a particular sense of a verb, participants use that sense's bias during language comprehension (Hare, McRae, and Elman, 2003). Similarly, constraints created by whether the initial noun in a sentence is a better cause, like "the brick", or a better theme, like "the glass", influence what sense participants expect verbs like "shatter" to have, and consequently what subcategorization bias they have (Hare, Elman, Tabaczynski, and McRae, 2009). These results suggest that adults maintain multiple fine-grained statistics about subcategorization biases for a single verb, which depend on the sense of the verb suggested by the sentence. Importantly, this suggests that biases are collected below the level of individual verbs, at an even narrower granularity than is typically contrasted in comprehension studies.

Biases are also governed by super-ordinate statistics about the behavior of entire classes of verbs. Examples of verb classes include dative verbs, which both tend to participate in similar structures, and to talk about similar kinds of transfer or communication events (Levin, 1993). Verb classes with relatively few verbs that alternate, or appear in two different syntactic structures, tend to promote learning of verb biases (Wonnacott, Tanenhaus, and Newport, 2008). Further, Wonnacott and colleagues found that the behavior of verb classes with many alternating items tends to reflect the general distributional properties of that class, rather than conforming to verb-specific statistics. Although it is more difficult to draw a direct parallel to an English verb class, the majority of transitive verbs can alternate between the passive and active, and may be a natural example of this type of class (e.g., Cureton, 1970). Based on the behavioral differences created by these two artificial verb classes, Wonnacott et al. conclude that these results show that learners are influenced by both the distributions of individual verbs and entire classes of verbs. Although this conclusion was derived from a study of artificial grammar learning, similar results have also been found with English verbs, demonstrating that when most verbs in a class alternate, children do not learn new verb biases (Lin, 2020). Beyond the behavior of verbs that participate in the same syntactic structures, verb biases are also influenced by whether verbs appear in related structures. Verbs like "spray" can appear in longer locative sentences ("She sprayed water onto the sidewalk"/ "She sprayed the sidewalk with water") as well as shorter transitive sentences ("She sprayed the sidewalk"/ "She sprayed the water") (Twomey, Chang, and Ambridge, 2014). Twomey and colleagues found that the transitive uses could be used to inform the locative biases of different verbs. Further, adults and nine-year-old children are able to use distributional cues like these to correctly generalize to new verbs (Twomey, Chang, and Ambridge, 2016). Together, these findings show that verb behavior is the result of multiple distributions. One of these distributions is a verb bias that is built up through clear verb-structure pairings, but other distributions may ultimately play a much larger role in how a verb behaves in usage.

Finally, distributional statistics and meaning interact with one another during learning to jointly determine how speakers generalize verbs. Like the statistical factors previously discussed, this interaction happens at both the level of the individual verb, and at the level of broader verb classes. At the level of individual verbs, usage seems to be influenced by both entrenchment and verb semantics. Entrenchment occurs when the frequent occurrence of a verb in a specific structure or structures discourages its use in other, unattested structures; for instance, seeing "disappear" only in intransitive sentences like "The rabbit disappeared" could prevent its use in other, incorrect structures (Braine and Brooks, 1995). By contrast, semantic classes of verbs are proposed to group verbs both broadly, such as "all dative verbs", and narrowly, such as "verbs specifying an instrument of communication", which allows verbs with similar meanings to generalize correctly to the same set of syntactic structures (Pinker, 1989).

Both entrenchment and meaning are required to explain how adults use verbs. When adults and children were asked to rate grammatical and ungrammatical sentences, all groups preferred grammatical sentences with high-frequency verbs like "laugh" more strongly than sentences with low-frequency verbs like "giggle" (e.g., Ambridge, Pine, Rowland, and Young, 2008; see also Ambridge, Pine, and Rowland, 2011). Additionally, grammaticality ratings are graded by verb semantics; ungrammatical sentences like "She laughed him" were rated lower than sentences like "She tumbled him" because laughter is more likely to be caused by an internal source (Ambridge et al., 2008). Extending novel verbs to these structures seems to be guided by relatively narrow semantic classes, such that novel "falling" verbs are constrained to intransitive structures in the same way as English verbs with similar meanings (Ambridge et al. 2011; see also Ambridge et al., 2008). Finally, the behavior of children in similar tasks suggests that adult representations of verb semantics and frequency are learned and refined over a relatively long developmental time course (Ambridge, Pine, Rowland, and Chang, 2012). Overall, these findings point to the conclusion that the usage-based statistics that create verb biases are likely the result of these same semantic and frequency-based constraints. Consequently, it is reasonable to ask whether subtle changes to semantic verb classes can also be induced by specific experiences, much in the way that previous studies of verb bias seem to actively influence the entrenchment of verb-structure combinations.

Meaning and statistics also compete with each other in determining syntactic structures, and sufficient competition from meaning may block the use of verb bias statistics. For example, novel verbs that alternated between two structures that both described transitive events acquired verb biases (Thothathiri and Rattinger, 2016). By contrast, when verbs alternated between structures that described either using an item as an instrument (tickle with a feather) or a modifier (the cat with a feather), Thothathiri and Rattinger found that participants described events with a particular structure rather than using a verb-specific bias. Similar results have been found when structures seem to be predicated by a discourse constraint, like the presence of a pronoun (Perek and Goldberg, 2015). However, it is not simply the case that meaningful structures totally block the use of verb bias statistics. While seeing a verb appear in only one class tends to make speakers less likely to generalize that verb to another structure, they will also still occasionally use verbs in the unattested structure when the meaning is appropriate (Perek and Goldberg, 2015; Perek and Goldberg, 2017). Additionally, when a verb is used in only one structure regardless of that structure's meaning, speakers learn not to generalize that verb, and tend to generalize other verbs less frequently (Perek and Goldberg, 2017). Finally, making a familiar English structure more strongly associated with a meaning instead of an individual verb causes speakers to rely on that meaning when choosing a structure to produce (Thothathiri and Braiuca, 2021). While verb biases are one part of explaining how structures are chosen during production, it is clear that the meaning of the structures themselves is also an important consideration.

To summarize, verb biases are not only the record of how often a verb co-occurs with a structure, but also the result of many other sources of distributional evidence. Additionally, this evidence is not restricted to pure co-occurrence statistics. Instead, evidence at all levels includes both meaning and statistical information, and these two sources of information actively compete with one another. Based on these findings, it is clear that verb biases are simply one aspect of a large, interrelated set of information sources about verb behavior. Some parts of this network are highly-studied; for instance, it is now well-established that children and adults can learn new verb biases by pairing particular verbs and structures (e.g., Coyle and Kaschak, 2008; Qi, Yuan, and Fisher, 2011; Ryskin, Qi, Duff, and Brown-Schmidt., 2017). Other, higher-order parts of this network have been less intensely investigated. While it is clear that meaning influences how verbs are used, and that the behavior of related verbs affects each other, whether training a specific verb affects semantically-related verbs is unknown. In other words, is updating a verb's bias constrained to that specific word? Or does it spread throughout this network, subsequently affecting verbs that behave in similar ways?

The following studies investigate these questions using both behavioral methods and a computational model. In the behavioral experiment, speakers are trained on dative and transitive verbs that occur in only one structure. For instance, the dative verb "toss" would only occur in the double object dative, as in "Tyler tossed his friend the ball." Similarly, a transitive verb, like "captivate", would only occur in the passive, as in "Jane was captivated by the ruins." At test, speakers are tested on both "toss" and a semantically-related verb, "throw." The test phase allows speakers to choose what structure to produce. For dative verbs, participants may produce a double object dative, or a prepositional dative, as in "Tyler threw the ball to his friend". For transitive verbs, participants may choose to produce a passive sentence, or they may produce an active sentence like “The ruins captivated Jane.” The free choice of structures allows participants to demonstrate their biases for both the trained and untrained verbs. If verb biases are specific to a particular verb, “toss” should show evidence of training, but “throw” should not. However, if information from verb bias training spreads to related verbs, then untrained verbs like “throw” should also show evidence of training. The computational model addresses whether the behavior seen in human participants can also be replicated using distributed semantic representations in a simple neural net.

3.1 METHODS

124 total participants were recruited from the University of Illinois at Urbana-Champaign participant pools. Participants received either $8 or 1 hour of course credit. Participants were native speakers of English with no reported language disorders. A total of 90 participants were included in final analysis, with participants being excluded for reasons given below.

Participants completed a training and a testing phase in the experiment, but were not told that the function of the phases was training and then testing. However, they were told that they would be asked to do two different sentence completion tasks in the two phases.

In the training phase of the experiment, participants completed sentence stems that were presented one at a time. For example, a participant could see a sentence like “The constant problems enraged ________”, and then would be asked to complete the stem. These sentence stems were intended to cause participants to complete the sentence with the structure associated with that verb. These sentence stems trained participants to acquire new biases four verbs, two of which were dative, and two of which were transitive. Each verb appeared in only one sentence structure; for instance, one dative verb always appeared in the double object dative, and the other always appeared in the prepositional dative. The combination of verbs seen by each participant was generated randomly. These lists were formed by selecting one of each of the following pairs of verbs: throw-toss, mail-ship, anger-enrage, and captivate-intrigue. So, an example of a possible training list might be “throw,” “ship,” “anger,” and “intrigue.” Participants completed 10 stems for each verb, for a total of 40 training trials. Dative and transitive trials alternated with one another.

In the testing phase of the experiment, participants were presented with a scrambled list of words, like “problems, Dave, enrage”, and were asked to create a sentence. Each scrambled list contained one of ten test verbs. Four were the verbs that the participants had been trained on in the first phase of the experiment, and four were untrained synonyms of those verbs. For example, if a person was trained on “throw”, they would be tested on both “throw” and “toss”. Finally, participants were also tested on one unrelated and untrained dative verb, either “give” or “hand”, and one unrelated and untrained transitive verb, either “impress” or “surprise”. These verbs were simply intended to act as fillers. For each trial, Participants completed 10 scrambled sentences for each verb, for a total of 100 test trials.

3.1.1 Coding

Responses were coded by research assistants using a pre-registered coding manual. The intent of this coding manual was to retain as much data and as many participants as possible, while still eliminating participants who failed to learn biases for any one of the trained verbs.

The 40 training trials were coded first. If these trials matched the structure that the verb was supposed to be trained to and made sense, then the response was accepted. Other responses, such as ungrammatical sentences or other syntactic structures, were rejected. If a participant had fewer than seven scoreable responses for any of the four training verbs, then their data was excluded from final analysis.

In the test phase, sentences were marked as accepted if they used either syntactic alternate for a particular verb. For instance, “toss” would be marked accepted regardless of whether it appeared in the double object dative or prepositional dative. Sentences that used any other syntactic structure, or which did not use the verb indicated, were rejected. Finally, sentences which had basically the correct structure, but which included some other phrases, were included in the final data. So, for an example test trial “Jim, ball, Janet, throw”, the sentence “Jim threw the ball away and made Janet mad” would be rejected, while the sentence “Jim threw the ball to Janet because they were friends” would be accepted, and marked as a prepositional dative. If fewer than 50% of the responses in the test phase were scoreable, then the entire participant was excluded from final analysis.

3.2 RESULTS

3.2.1 Comparisons of Trained and Untrained Verbs

Results from the dative and transitive groups were analyzed independently, and were further separated based on whether the verbs in each group were trained or untrained. Results from both groups of dative verbs are plotted in Fig. 10 below. Comparisons were done using a Wilcoxon signed-rank test. Verbs that were trained to appear in the double object dative were more likely to do so than verbs that were trained to appear in the prepositional dative (DO-trained mean: 0.556, PD-trained mean: 0.458, $W=896.5$, $p<0.05$). Similarly, the untrained synonymous verbs showed transfer of this effect. Verbs with DO-trained synonyms were more likely to occur in the double object dative than verbs with PD-trained synonyms (DO-untrained mean: 0.518, PD-untrained mean: 0.446, W=1282.5, p<0.05).

However, the transitive verbs show effects of neither training nor testing. The verbs that were trained in the passive were not more likely to appear in that structure than verbs that were trained to appear in the active (Passive-trained mean: 0.426, Active-trained mean: 0.350, W=1409.5, p>0.05). Similarly, there was no effect of training on synonyms of these verbs (Passive-untrained mean: 0.395, Active-trained mean: 0.435, W=1488.5, p>0.05).

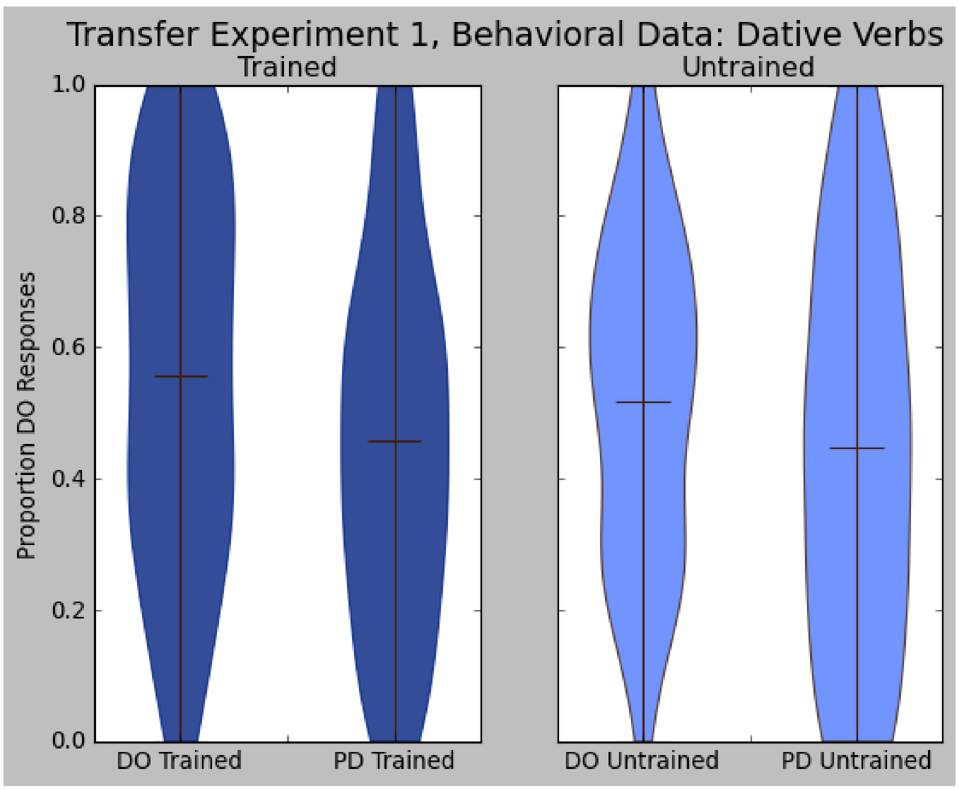

Figure 10: Plots for dative verbs (trained and untrained).

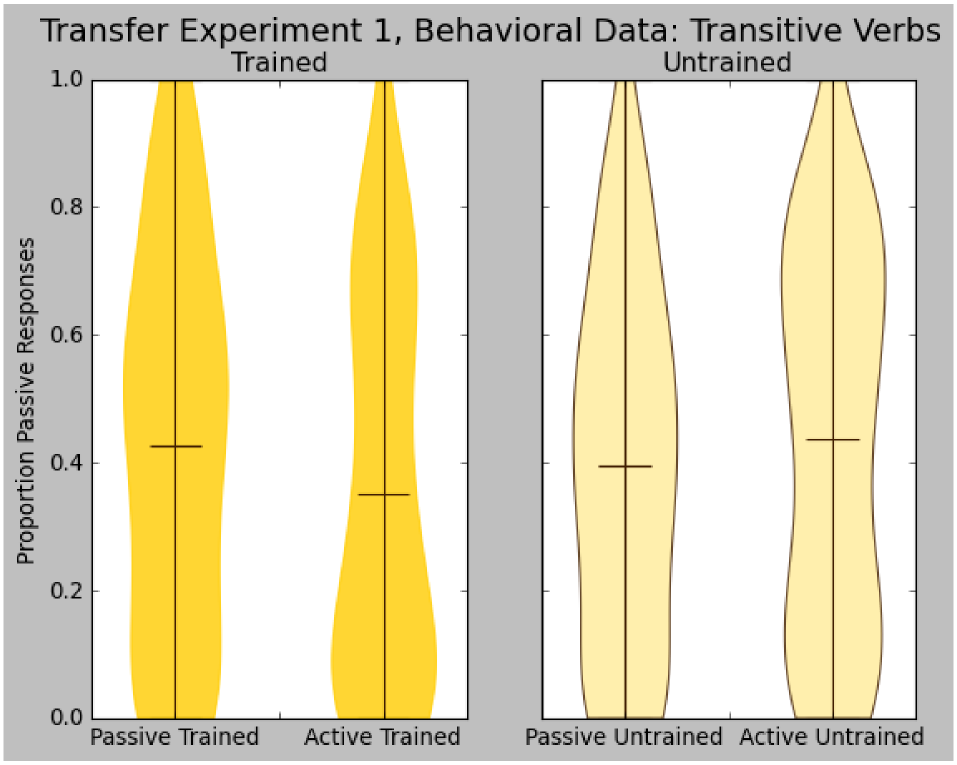

Figure 11: Plots for transitive verbs (trained and untrained).

3.2.2 Comparing First and Second Halves of Blocks

In order to better understand how production changed over the course of the block, the differences between the first and second half of each block was compared. The purpose of this comparison was to determine whether the training effect was larger in the first half of the block for any of the conditions, or if the effect persists throughout.

For the trained and untrained dative verbs, this involved finding the difference between the DO trained and PD trained verbs. For the trained verbs, two participants were dropped because they were missing means in one of the halves of the blocks. The difference in the size of the training effect between the first half and the second half of the block was significant, although the effect was not particularly large (First Half: 0.14, Second Half: 0.04, W=1048.5, $p<0.05$). For the untrained verbs, seven participants were dropped because they lacked a mean in one of the conditions. After this, the difference between the remaining 83 participants in the first and last halves of the untrained verbs was not significant (First Half: 0.12, Second Half: 0.04, W=1328.0, p>0.05). However, it is important to note that both the trained and untrained verbs have a numerically larger training effect in the first part of the block regardless of significance.

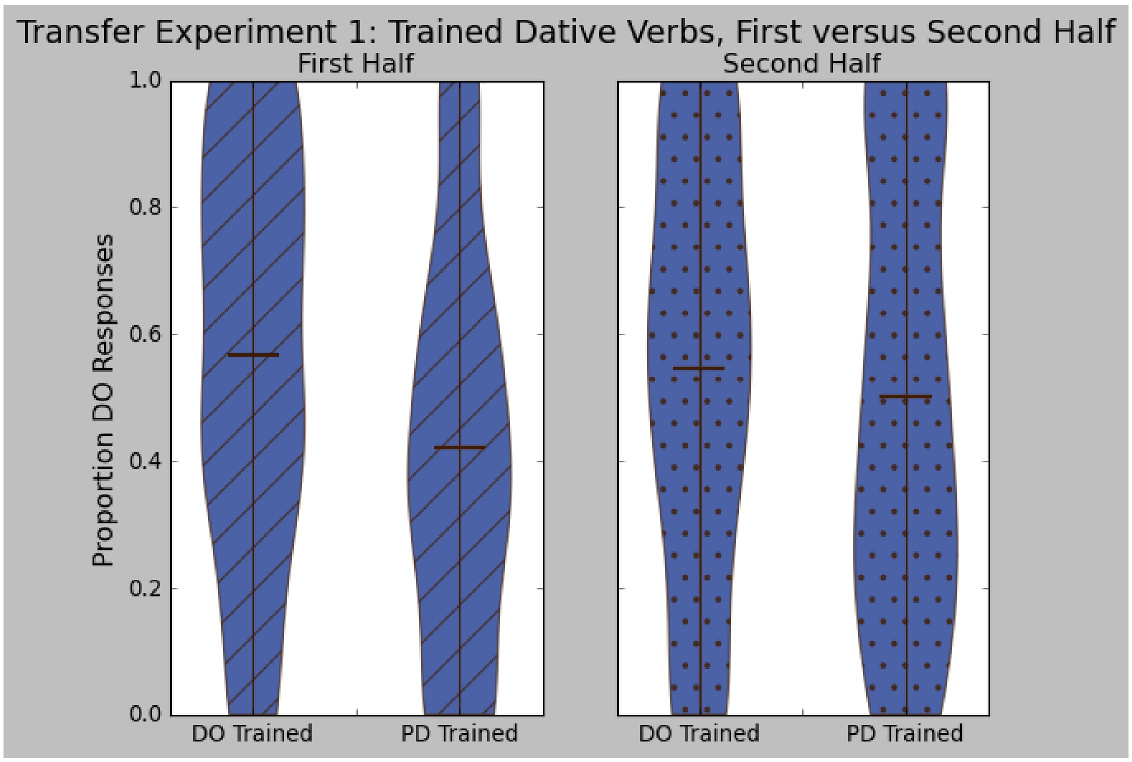

Figure 12: Proportion DO responses for the first half and second half of the trained dative verbs, divided by block.

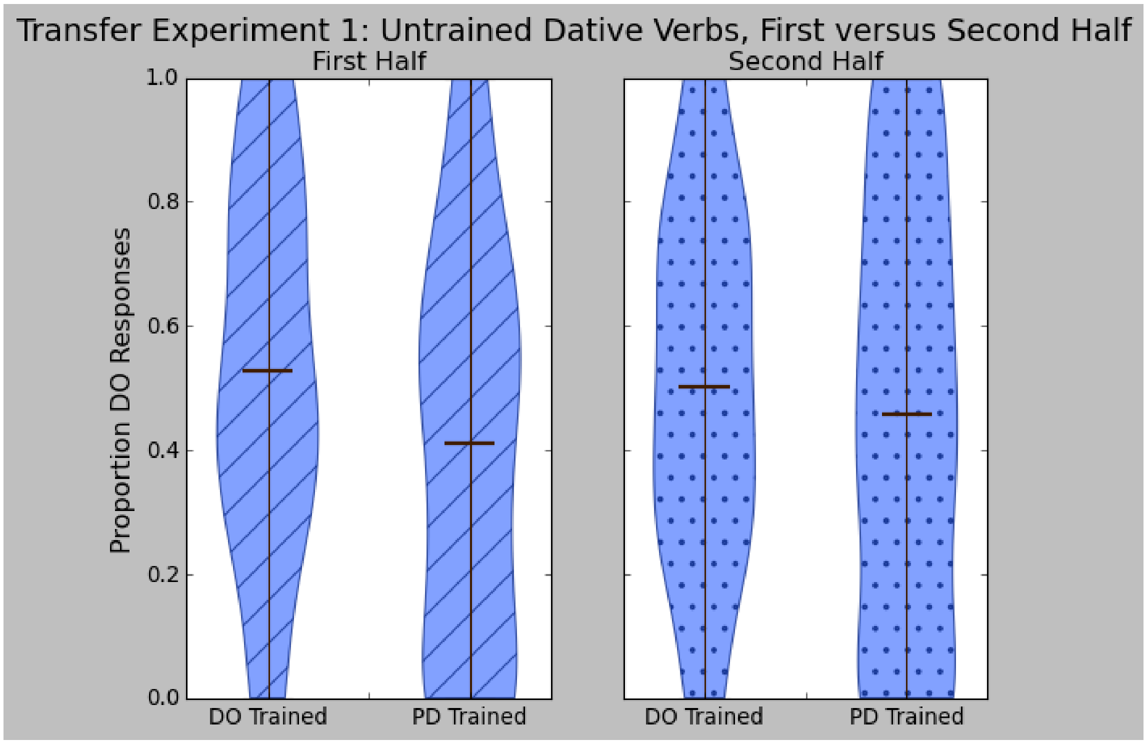

Figure 13: Proportion DO responses for the first half and second half of the untrained dative verbs, divided by block.

When the transitive verbs were divided between the first and second half of the testing block, they followed a similar pattern. The difference in the size of the training effect was significant between the first and the second half of the trained transitive verbs, although the effect was again rather small (First Half: 0.11, Second Half: 0.03, W=1142.0, $p<0.05$). Finally, the difference in the training effect of the untrained verbs was not significant (First Half: -0.07, Second Half: 0.02, W=1287.0, $p>0.05$).

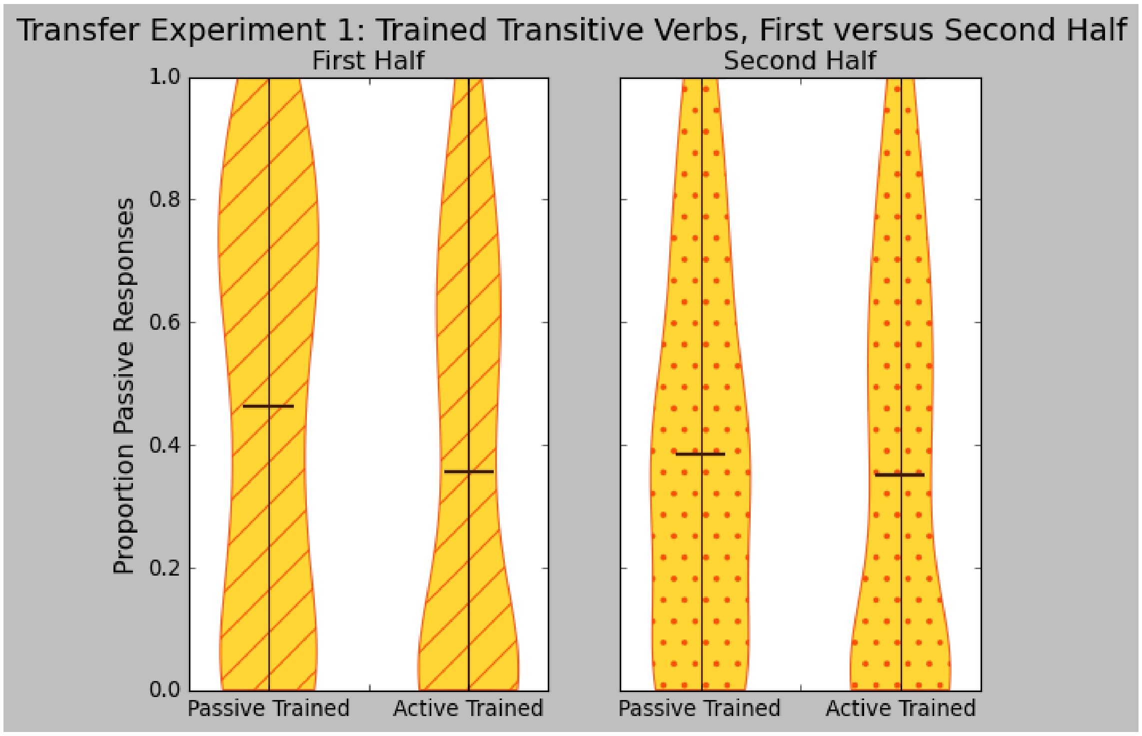

Figure 14: Difference in the size of the training effect between the first and second half of the trained transitive verbs.

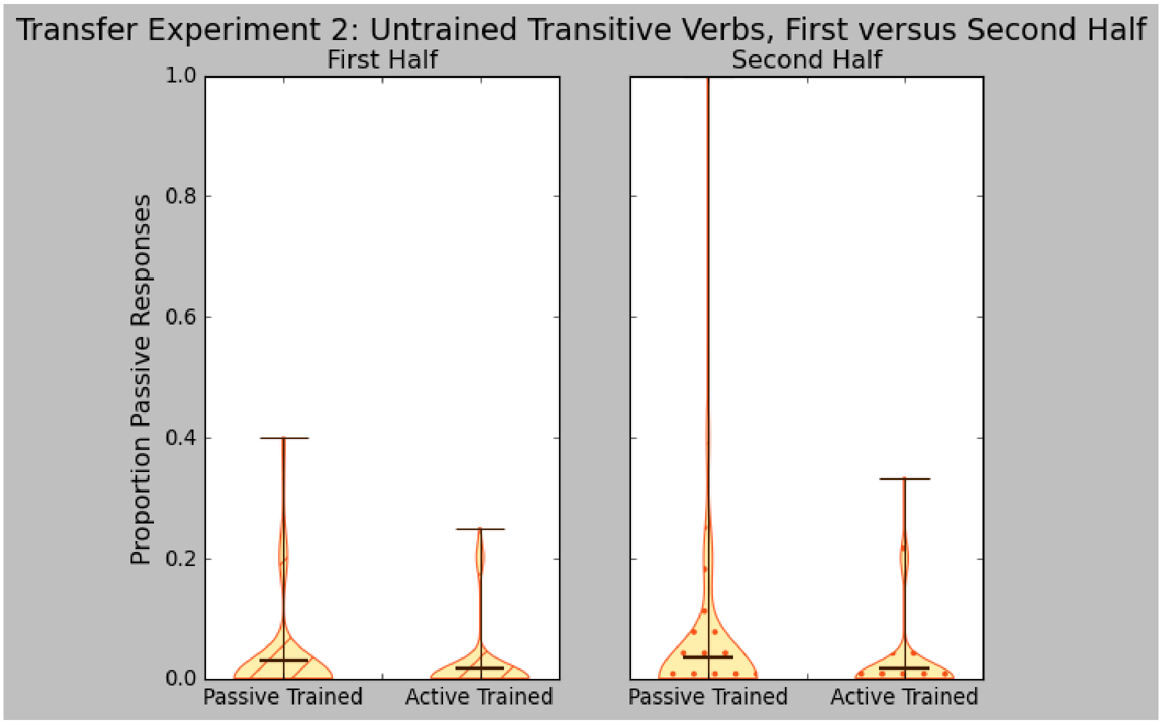

Figure 15: Difference in the size of the training effect between the first and second half of the untrained transitive verbs.

3.3 CONCLUSIONS

This study suggests two primary points and one corollary. First, learning transfers between semantically-related dative verbs. Second, learning does not transfer between semantically-related transitive verbs, and these verbs are remarkably resistant to acquiring new verb biases. These points contribute to better understanding the learning mechanism that updates verb biases, and suggest constraints on future modeling projects. Finally, the follow-up analyses of the size of the training effect at the beginning and end of the block suggest that learning continues throughout the testing block. Implications for further empirical work are discussed below.

The first finding is that learning seems to transfer from verbs with newly-acquired verb biases to synonymous verbs. This process is certainly not the primary way that verb biases are learned, since there are some verbs (e.g., give and donate) with similar meanings that maintain separate verb biases. However, transfer between similar verbs might belong to a class of statistics that partially-constrain the distributions of verbs. These include statistics like distributions of the same verb in related structures. While not direct evidence, both models and humans are able to use distributions like these to correctly constrain the behavior of verbs in classes like the locative (Twomey, Chang, and Ambridge, 2014; Twomey, Chang, and Ambridge, 2016). Additionally, learning that spreads from one verb to another would likely be overwhelmed by any direct experience with that verb in a structure. Ideally, transfer between synonymous verbs would serve as a nudge in the right direction; at worst, the learning would be disregarded. While not as central to verb bias learning as direct experience with verb-structure co-occurrences, transfer is probably one of many statistics that act as a weak constraint on learning.

The second major finding is a replication of the lack of a verb bias training effect for transitive verbs, despite a numerical trend in the right direction for the trained transitive verb. Similar findings are also reported in Kelley (2019), which twice found a lack of a training effect for transitive verbs. Replicating this finding with nearly twice as many participants as in either of the Kelley (2019) studies suggests that the lack of an effect is not due to insufficient power or some specific aspect of those experiments. Additionally, participants in the transfer experiment produced passive structures approximately 40% of the time, suggesting that a possible training effect is not obscured by a floor effect. Although it is not definitive, the most likely explanation is that there is previous learning about the behavior of passives that blocks the acquisition of a new verb bias. Previous work suggests that behavior at the level of an entire class of verbs can block learning about the biases of individual verbs (Wonnacott, Newport, and Tanenhaus, 2008). Potentially, the behavior of transitive verbs as a group blocks the learning of individual verb biases because there are other cues that are more predictive of an upcoming structure. Humans can learn to shift which cues they use to predict structure in an experimental setting (Thothathiri and Braiuca, 2021). Consequently, in an experiment that was not designed to shift those cues, it is possible that any learning about verb biases was blocked by the significant amounts of life experience that participants have with these verbs. The lack of a verb bias training finding for transitive verbs is likely the result of learning dynamics in the language production system rather than a failure to find an effect.

Overall, this experiment presents an interesting divide between dative and transitive verbs. The dative verbs show both training and transfer effects, and the transitive verbs show neither. There are compelling reasons to think that these effects represent real aspects of verb bias learning, but replicating these effects is also important. First, because these effects may depend on the behavior of individual verbs, it is important to replicate these findings with a different set of verbs. Second, the verbs in this study were chosen to be similar, but this similarity was not normed objectively. Consequently, it is also important to both determine the similarity of the verbs used in this study, and to select verbs for future studies in a more deliberate manner.

Finally, the change in the size of the training effect between the beginning and end of the blocks has implications for future work. For both sets of trained verbs, the difference at the beginning of the block was larger than the difference in the second half, although in both cases these effects were small. However, assuming that this effect is real, then this has implications for optimizing the design of future empirical verb bias training work. One interpretation of these results is that they are evidence that producing other related structures during the testing block is blunting the effect of training, which may be evidence of self-priming (e.g., Jacobs, Cho, and Watson, 2019). Consequently, future verb bias training experiments might want to limit the number of testing trials, and achieve sufficient statistical power by adding participants instead of trials. However, given the small size of this effect, any conclusions about theoretical implications or future experimental design are somewhat limited.

CHAPTER 4: TRANSFER NORMING AND REPLICATION

The results of the initial study suggested that semantic information transfers between semantically-similar dative verbs, but not between semantically-similar transitive verbs. This finding can potentially be explained in a number of ways. Some of these explanations have important theoretical implications. For example, these results could imply that activation is passed differently between dative verbs than it is between transitive verbs, which would suggest there are fundamental differences in their semantic networks. Other explanations are less theoretically-relevant; for instance, dative verbs could be more concrete than transitive verbs. Given the small number of verbs used in the previous study, it is difficult to differentiate between these explanations. Consequently, generalizing these results to other verbs is crucial to understanding these results.

In order to choose new verbs, a norming study was performed to determine the concreteness and similarity of various semantically-similar pairs of dative and transitive verbs. Initially, the intent was to choose verbs that varied in concreteness, while controlling for levels of similarity. An initial study found that identifying similar dative verb pairs was difficult, and that it was particularly difficult to identify similar abstract dative verb pairs. A follow-up study was used to identify a second concrete dative verb pair, and to collect similarity and concreteness judgments for the verbs used in the original study.

4.1 NORMING STUDY 1

4.1.1 Methods

A total of 30 participants were recruited from the University of Illinois Course Credit pool, and were compensated by receiving 0.5 course credits. Three participants failed to respond to attention checks, which requested that participants respond with a particular number on a rating scale. Consequently, these participants were not included.

Participants were first asked to rate the concreteness of all the individual verbs included in the study. Participants were offered a brief explanation of concreteness, which was defined as verbs that were easy to physically sense. Participants were also told that the opposite of concreteness is abstractness, and that abstractness refers to verbs that are difficult to sense. Finally, participants were instructed to use the sense of the verb presented in a context sentence, rather than using the sense of the verb that might first come to mind. Participants responded on a scale that scale that included six points, ranging from 5 (Very concrete) to 1 (Very abstract), and an additional sixth point that stated that the participant did not know that word. The verbs rated in this study are presented in Table 3.

Participants were then asked to rate the similarity of pairs of verbs. As with concreteness, similarity was defined for the participants. Specifically, it was defined as verbs that are used in similar contexts and mean similar things. Again, participants were instructed to use the senses of the verbs presented in context sentences to generate their similarity ratings. Participants responded on a scale that included five points, ranging from 5 (Almost always) to 1 (Almost never). The pairs of verbs rated in this study are also indicated in Table 3.

The study was run using jsPsych (de Leeuw, 2015). Data was cleaned using Python, and analyzed using R (R Core Team, 2019). In addition to the three participants who were eliminated for failing an attention check, individual ratings from participants were removed if participants indicated that they did not know particular verb. For example, if participants indicated that they did not know the word “relay”, both the concreteness rating for “relay” and the similarity rating for “relay/radio” would be removed. In total, five concreteness ratings and four similarity ratings were removed from final analysis. Finally, if participants did not respond to a particular question, that line would also be removed. A total of three concreteness ratings and one similarity rating were removed for lack of response.

4.1.2 Results

Table 3: Norming Experiment 1 Results

Verb	Transitivity	Concreteness	Concreteness Rating	Similarity	Context Sentence
Will	Dative	Abstract	1.88	2.65	Henry willed the house to his daughter.
Leave	Dative	Abstract	3.63		Charlotte left the house to her niece.
Read	Dative	Abstract	3.74	2.85	Diane read the sign to Tim.
Tell	Dative	Abstract	3.56		Vera told a story to Natalie.
Radio	Dative	Abstract	3.48	3.04	The security guard relayed the message to his partner.
Relay	Dative	Abstract	2.71		The soldier relayed the message to her commander.
Hand	Dative	Concrete	4.27	4.26	Jennie handed the papers to Trent.
Pass	Dative	Concrete	4.33		Dorothy passed the cup to Frank.
Tug	Dative	Concrete	4.33	4.04	Karen tugged the heavy box to the librarian.
Drag	Dative	Concrete	4.46		Wayne dragged the firewood to his friends.
Carry	Dative	Concrete	4.52	2.96	Flora carried the books to Gerald.
Take	Dative	Concrete	4.11		Matt took the coffee to his boss.
Doubt	Transitive	Abstract	1.7	3.93	Don doubted Joe's integrity.
Distrust	Transitive	Abstract	1.54		Ellie distrusted the claims.
Study	Transitive	Abstract	2.93	3.78	Tom studied the results.
Analyze	Transitive	Abstract	2.04		Asha analyzed the book.
Recall	Transitive	Abstract	2	4.78	Paul recalled the address.
Remember	Transitive	Abstract	1.85		Dani remembered the information.
Slice	Transitive	Concrete	4.59	4.04	Chris sliced the carrots.
Chop	Transitive	Concrete	4.3		Megan chopped the garlic.
Hit	Transitive	Concrete	4.56	4.04	The bicyclist hit the tree.
Strike	Transitive	Concrete	4.04		The driver struck the railing.
Push	Transitive	Concrete	4.44	4.3	Erica pushed the couch.
Shove	Transitive	Concrete	4.48		Tamara shoved the boxes.

Full reporting of all verbs can be seen in Table 3. Based on these results, one concrete dative verb pair (hand/pass) and two concrete transitive verb pairs (slice/chop and push/shove) were identified as being both similar to each other and highly concrete. While the abstract transitive verbs were rated as similar to one another, the ratings of the abstract dative verb pairs are not as similar as the other types of verb pairs. Finally, while “tug/drag” was rated as highly similar, these verbs do not typically alternate and were consequently excluded.

4.1.3 Conclusions

Based on the ratings in the first study, “hand/pass”, “slice/chop”, and “push/shove” were selected as verb pairs for the replication of the transfer experiment. Due to the lower similarity ratings of the abstract dative verbs, varying the abstractness of the verbs was abandoned as a manipulation. In order to obtain a second concrete verb pair that alternated relatively frequently, a second norming study was run. This study also included verb pairs from the first transfer study to allow comparison between studies.

4.2 NORMING STUDY 2

4.2.1 Methods

Methods for the second norming study were nearly identical to those used in Study 1. As in the first study, 30 participants were recruited from the University of Illinois Course Credit pool, and were compensated with 0.5 course credits for completing the norming study. Two participants failed attention checks that required them to respond with particular numbers on the rating scales. Consequently, these participants were excluded from final analysis.

Methods in the second norming study were identical to those in the first norming study, except that a smaller set of verbs were used. Participants were asked to rate the concreteness and similarity of the verbs in the study. Again, the first part of the study asked participants to rate the concreteness of each individual verb, and the second part asked them to rate the similarity of pairs of verbs. The same definitions of concreteness and similarity were used from the first study, and participants rated the verbs using the same scales. The experiment was presented using jsPsych (de Leeuw, 2015), and data was cleaned using Python and analyzed using RStudio. Additionally, ratings were removed from final analysis if participants reported that they did not know a verb, as described above. In total, one concreteness trial was removed because the participant reported that they did not know the verb, and one similarity rating was consequently removed as well. Finally, one similarity rating was removed because a participant did not respond.

4.2.2 Results

Table 4: Norming Experiment 2 Results

Verb	Transitivity	Concreteness Rating	Similarity	Context Sentence
Throw	Dative	4.57	4.11	Diane threw the ball to Jake.
Toss	Dative	4.5		Allie tossed an apple to Dave.
Ship	Dative	4.04	4.48	Victor shipped the papers to Fred.
Mail	Dative	4.46		Blanche mailed the card to Frank.
Take	Dative	3.89	3.19	Evan took the book to Max.
Bring	Dative	4.11		Greg brought the pillow to Erin.
Anger	Transitive	3.07	4.5	The team angered the coach.
Enrage	Transitive	3.07		Mary enraged her teacher.
Intrigue	Transitive	1.64	3.79	The forest intrigued Sam.
Captivate	Transitive	1.75		Emily captivated Megan.

Full reporting of all verbs can be seen in Table 4. Generally, the verbs presented in the first experiment are relatively similar to each other. Additionally, the dative verbs are rated as highly concrete, while the transitive verbs tend to be more abstract. As a pair, take/bring are less similar than other verbs in this experiment. However, both are concrete alternating dative verbs, and consequently they were selected as a pair for the replication.

4.2.3 Conclusions

This norming study was intended to select relatively similar verb pairs for a replication of the transfer study. These verbs needed to alternate between two structures, be similar to each other, and be concrete. Based on these constraints, four verb pairs were selected: “hand/pass”, “take/bring”, “slice/chop”, and “push/shove”. Additionally, this norming study determined the similarity of the four verb pairs used in the original transfer experiment: “ship/mail”, "throw/toss", "anger/enrage", and "captivate/intrigue". Although the verbs selected for this study represent a range of similarities, norming these verbs allows them to be compared post hoc in order to determine whether qualities of these verbs resulted in differences between the original and replication experiments.

4.3 TRANSFER REPLICATION

In order to fully understand the results of the previous transfer study, these new verbs were used for a replication. While this study used different verbs, the methods were nearly identical to those described in the first transfer study. The methods section below notes adjustments that were made in order to accommodate running the replication study online.

4.3.1 Methods

A total of 182 participants were recruited from the University of Illinois at Urbana-Champaign Course Credit Subject Pool. These participants were recruited for an online study, and were automatically credited with 1 hour of course credit for completing the study. Participants identified as native speakers of English. As in the first transfer experiment, participants were excluded if they had fewer than seven scoreable responses per verb, or if fewer than 50% of their testing data was useable. After applying these criteria, 88 participants were included in final analysis.

As in the previous replication experiment, participants learned a new verb bias for one half of a synonym pair, and then were tested on both the trained verb and the untrained synonymous verb. For this replication, the dative verb pairs were "hand/pass" and "take/bring", and the transitive verb pairs were "shove/push" and "chop/slice". As in the first study, participants saw ten training trials for each verb, receiving a total of forty training trials. As before, verbs were trained toward a specific syntactic alternate; for instance, “hand” would be trained toward the prepositional dative, while “take” would be trained toward the double object dative. Consequently, participants saw an equal number of each structure during training, but would demonstrate their new verb bias based on the structure they chose to use with that verb. Dative and transitive training trials were interleaved. Unlike the previous experiment, this study was run using a Qualtrics survey. Qualtrics does not allow for the same type of randomization as Psychopy. Consequently, training was presented in a total of eight lists, which are listed in the table below:

Table 5: List Structure for Transfer Replication

List	**PD**	**DO**	**Active**	**Passive**
1	Pass	Bring	Push	Slice
2	Hand	Take	Shove	Chop
3	Take	Pass	Chop	Push
4	Bring	Hand	Slice	Shove
5	Pass	Bring	Shove	Chop
6	Hand	Take	Push	Slice
7	Take	Pass	Slice	Shove
8	Bring	Hand	Chop	Push

As in the first transfer study, testing trials presented a list of scrambled words such as “hand, ribbon, Polly, Peter,” and participants were asked to use these words to create a sentence. Participants saw ten trials each of the four trained verbs and the four untrained verbs, as well as ten trials each of two untrained, unrelated verbs, for a total of one hundred testing trials. The untrained, unrelated verbs were “give” and “scare.” Because dative and transitive trials were interleaved and Qualtrics cannot generate this type of list structure, test trials were presented in eight pseudo-random lists.

Finally, during the process of data collection, it was particularly difficult to elicit complete training sets for lists that trained “take” to appear in the double object dative. As in the first experiment, participants were asked to complete stems such as “Ben took the doctor _______.” “Take” tends to encourage endings that are not a dative (“to a party”) or that are not grammatical (“stethoscope”). Exclusion rates for lists including “take” were extremely high, leading to concerns that continuing the experiment as planned would create selection effects. Consequently, articles were added to the stems for the last 20 participants of the experiment. For example, participants would now see stems like “Ben took the doctor the _________,” which eliminated many ungrammatical endings, while still allowing for some freedom of choice (e.g., answers like “the long way home”).

Coding proceeded in the same way as described in the first transfer experiment.

4.3.2 Results

Data were cleaned using the pandas package for Python (The Pandas Team, 2020; McKinney, 2010).

Testing trials were divided based on whether they used transitive or dative verbs, and these two groups were further divided based on whether the verbs were trained or untrained. When dative verbs that were trained toward a new double object dative bias were more likely to be produced in the double object dative than verbs that were trained toward the prepositional dative (DO-trained mean: 0.56, PD-trained mean: 0.42, $W=963.0$ $p<0.05$). However, there was no significant difference between dative verbs with synonyms that were trained toward a new verb bias (Synonym DO-trained mean: 0.54, PD-trained mean: 0.51, $W=1011.0$, $p>0.05$). Additionally, neither trained nor untrained transitive verbs showed a significant difference in the proportion of passive structures between verbs that were trained toward the passive or the active (Passive-trained mean: 0.03, active-trained mean: 0.01, $W=179.5$, $p>0.05$; synonym passive-trained mean: 0.03, synonym active-trained mean: 0.02, $W=105.0$, $p>0.05$).

While the training effectively changed the verb biases of the dative verbs, this training did not transfer to the untrained dative synonyms. Additionally, there is no evidence that training transitive verbs changed their verb biases, and consequently there is no evidence that this learning transferred to the synonymous verbs. However, the results from the transitive verbs are tempered somewhat by the extremely low proportion of passive trials. Finally, participants were qualitatively less likely to produce dative responses with the verb “take.” Although these observations do not change the results reported above, their implications for the differences between the first and second transfer experiment will be discussed further in the conclusions.

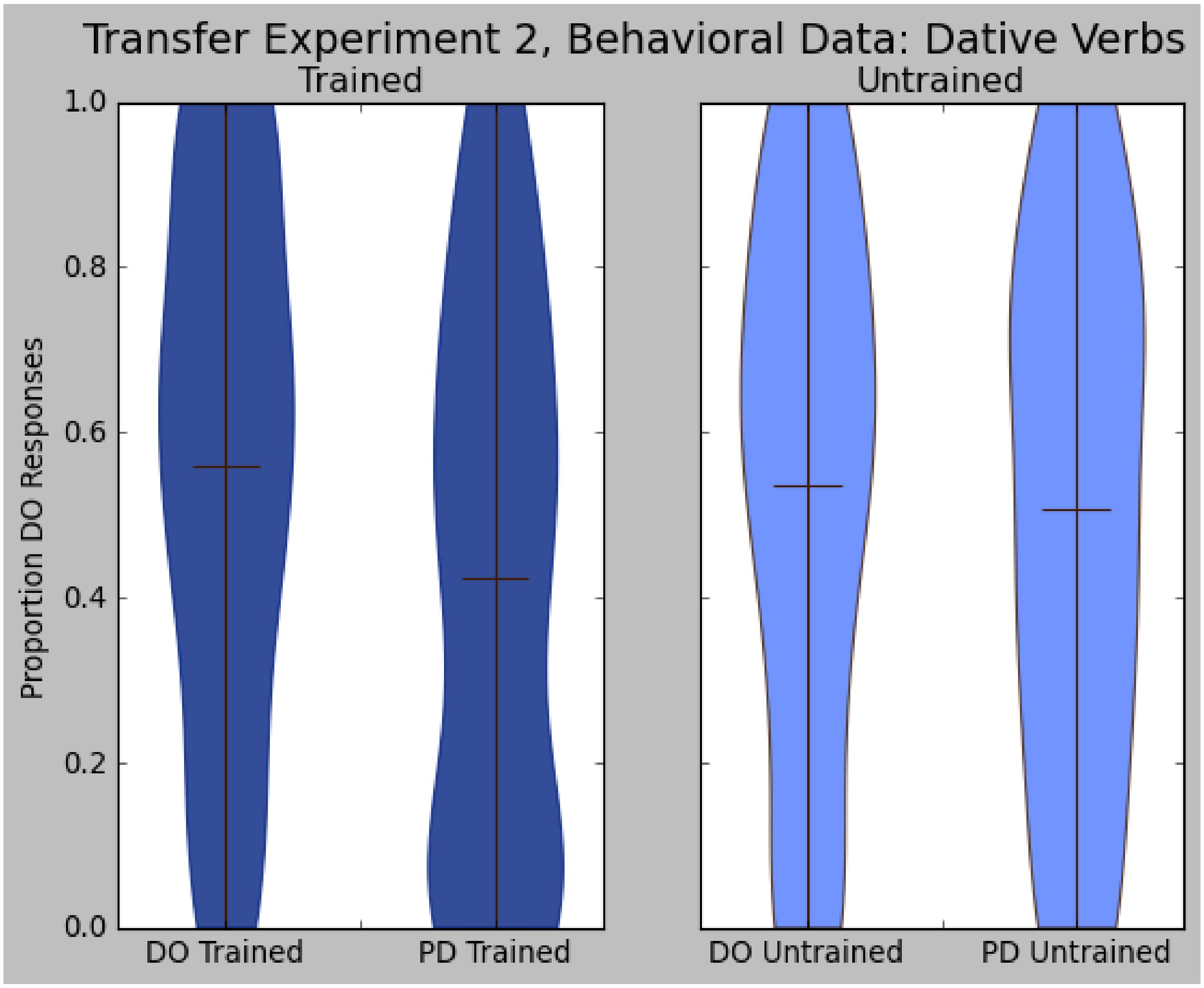

Figure 16: Trained and untrained dative verb results from transfer replication

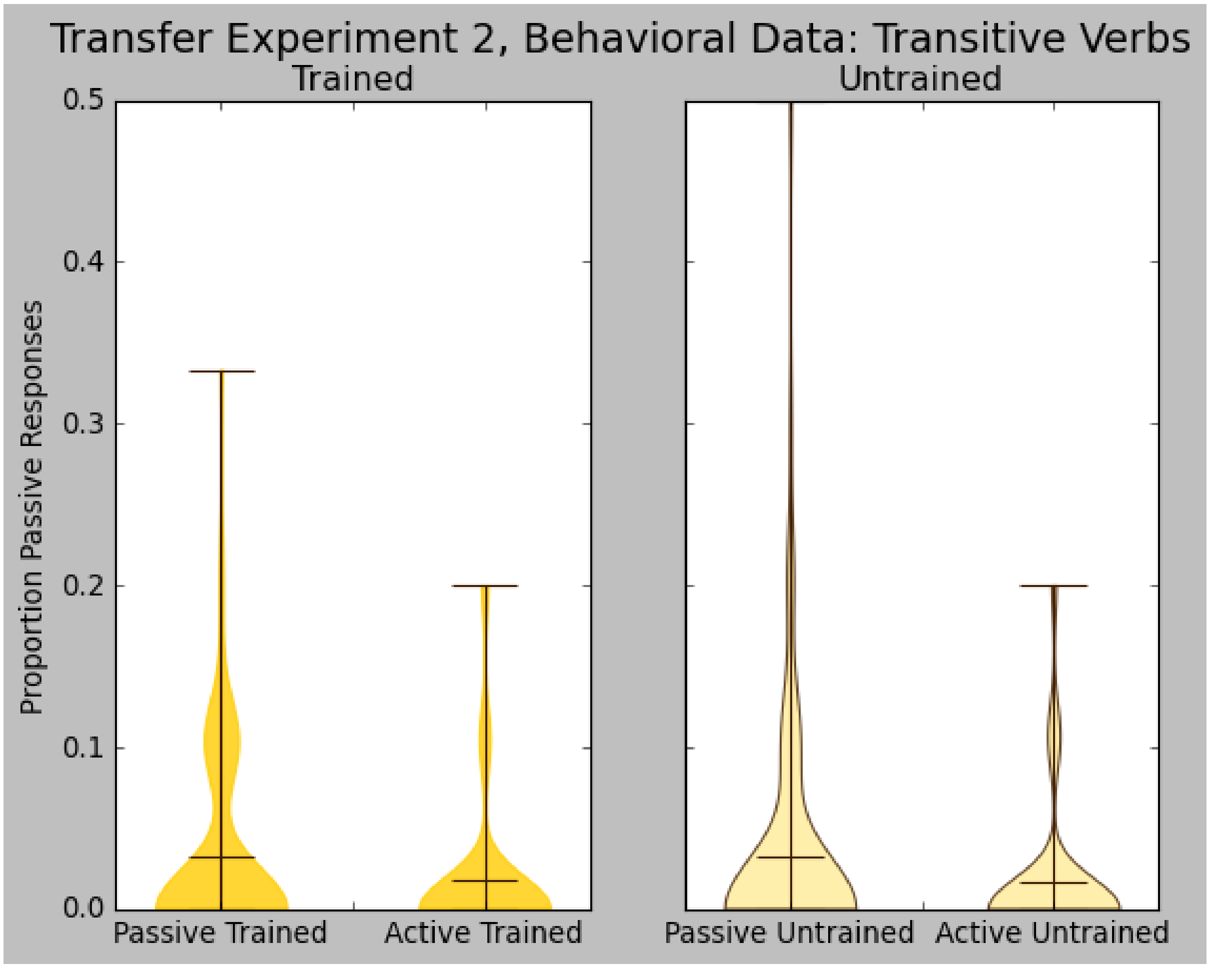

Figure 17: Trained and untrained transitive verbs from transfer replication.

Finally, in order to follow up on the analyses of the first and second half of the blocks from Chapter 3, the same analyses were completed on the replication data. Again, the difference between training conditions in the first half of the block was compared to the same difference in the second half of the block for both dative and transitive verbs. For the trained dative verbs, 21 participants were dropped to ensure that all participants had means for each cell. Analyzing the remaining data, no difference was found in when comparing the training effect between the first half and the second half of the block (First Half: 0.12, Second Half: 0.16, W=886, p>0.05). Twenty-seven participants were dropped from the untrained dative verbs so that every participant would have a mean for every training condition and both halves of each block. After this, no difference in the size of the training effect was found when comparing the first half of the untrained datives with the second half (First Half: 0.03, Second Half: 0.06, W=702, p>0.05). Results for both are presented in Figures 18 and 19.

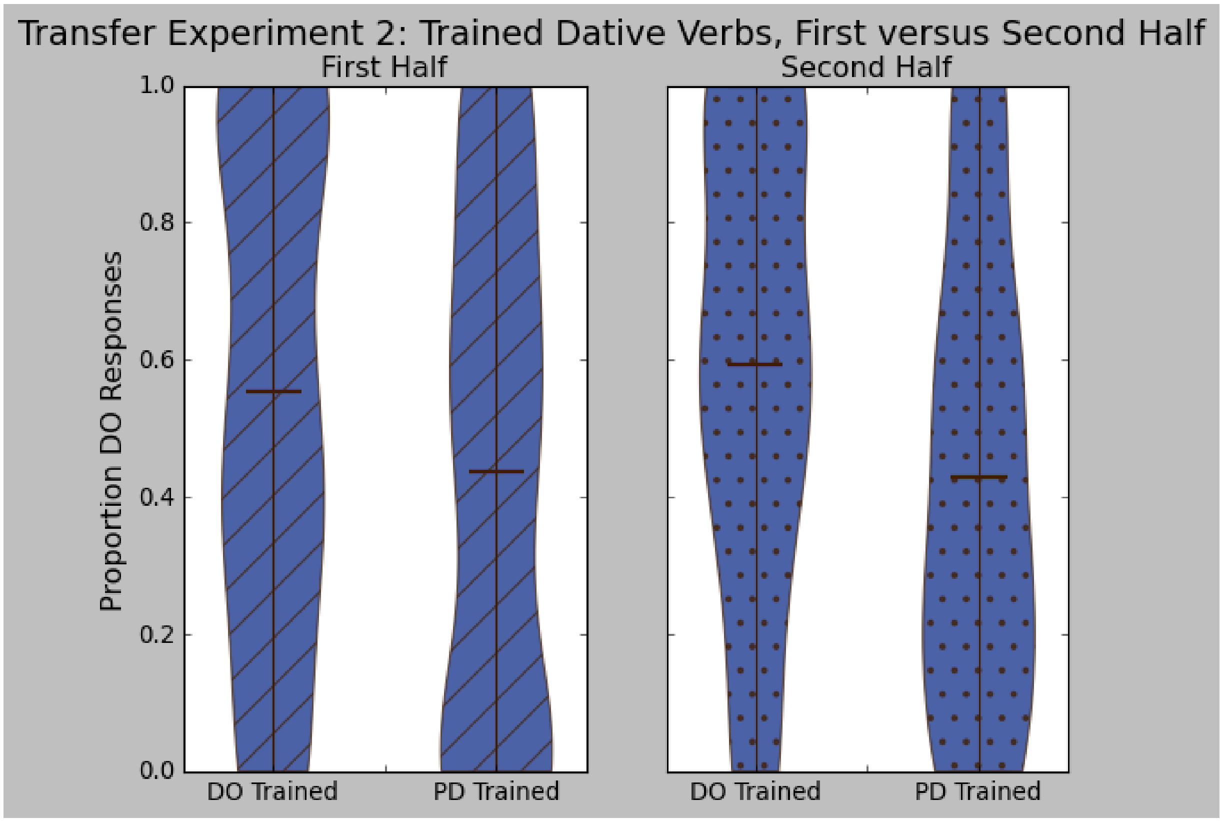

Figure 18: Transfer effect for trained dative verbs, split by half of block.

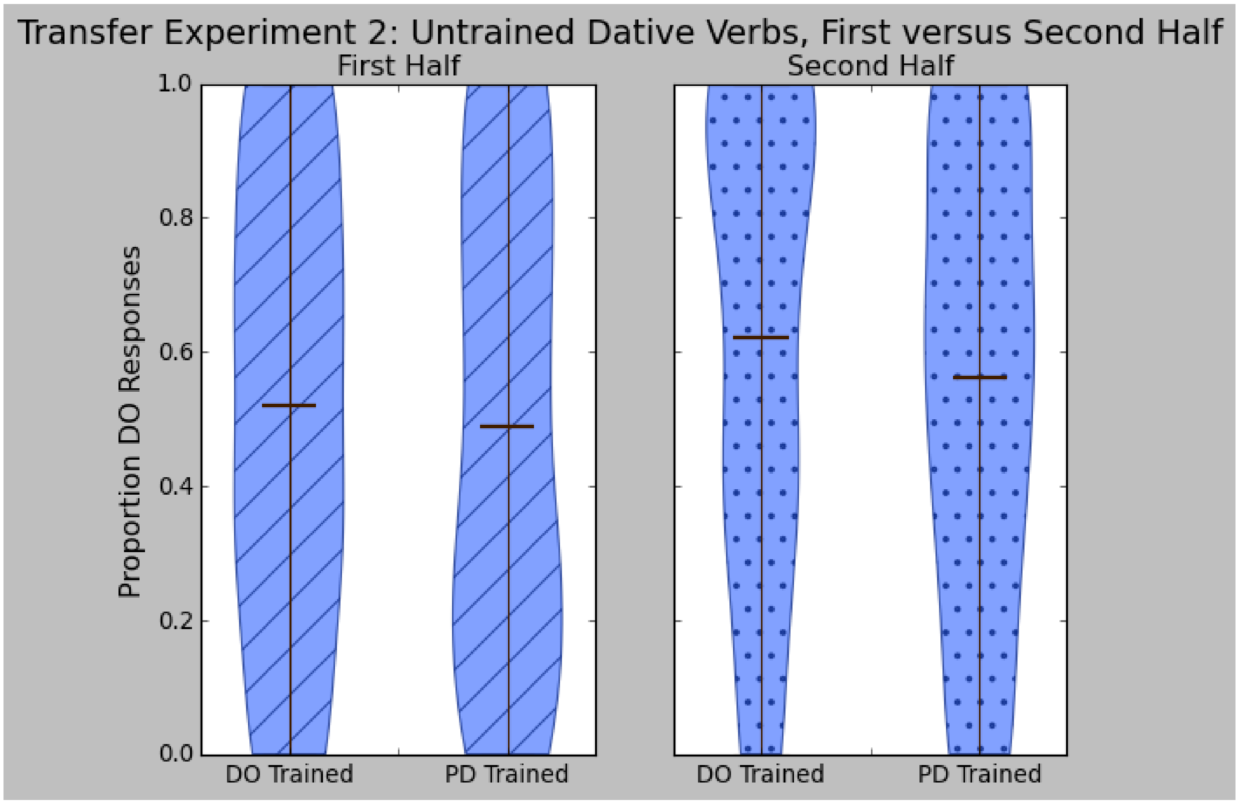

Figure 19: Transfer effect for untrained dative verbs, split by half of block.

This analysis was also run for the transitive verbs, although given the size of the effect overall it would be difficult to detect a change. Comparing the halves of the trained dative verb blocks, the size of the training effect is not different (First Half: 0.02, Second Half: 0.01, W=179, p>0.05). The two halves of the untrained transitive block also showed no difference in the size of the training effect (First Half: 0.01, Second Half: 0.02, W=92, p>0.05). Training effects for both sets of verbs and for both halves of the block are visualized below in Figures 20 and 21.

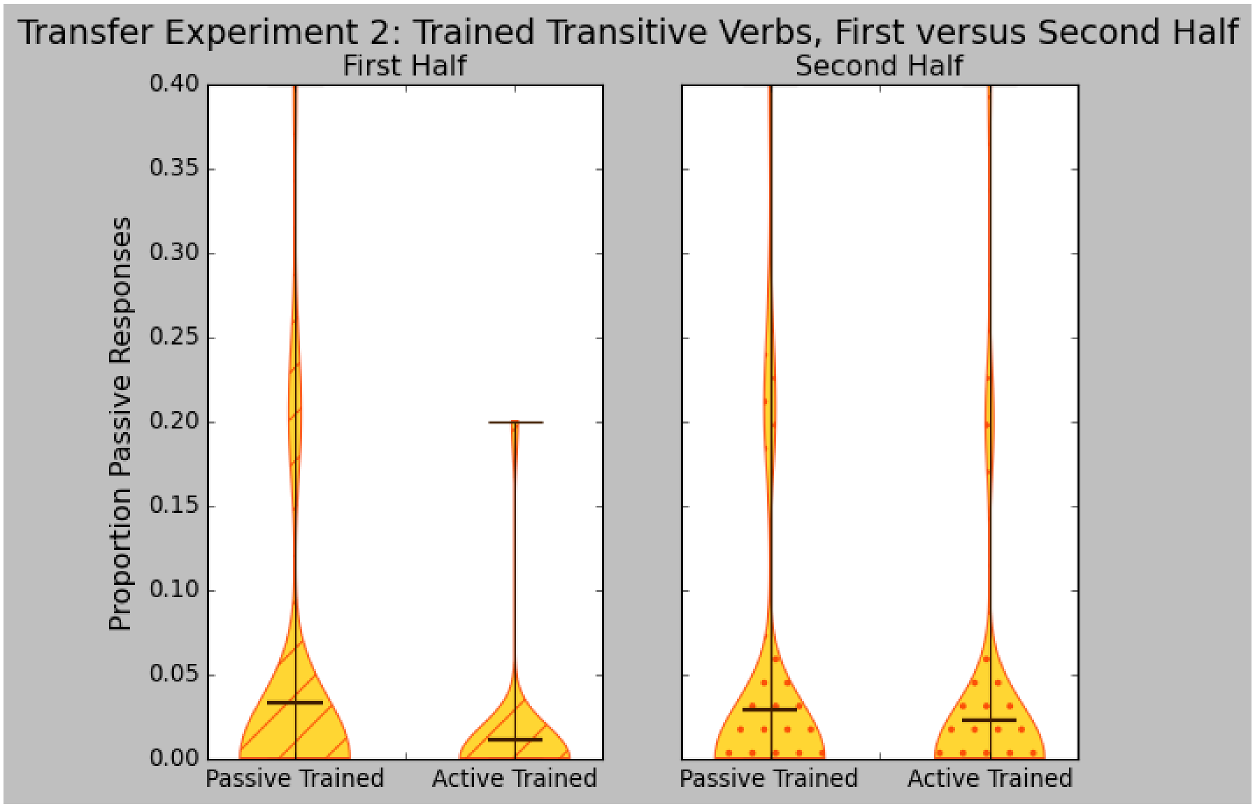

Figure 20: Training effect for trained transitive verbs, split by half of block.

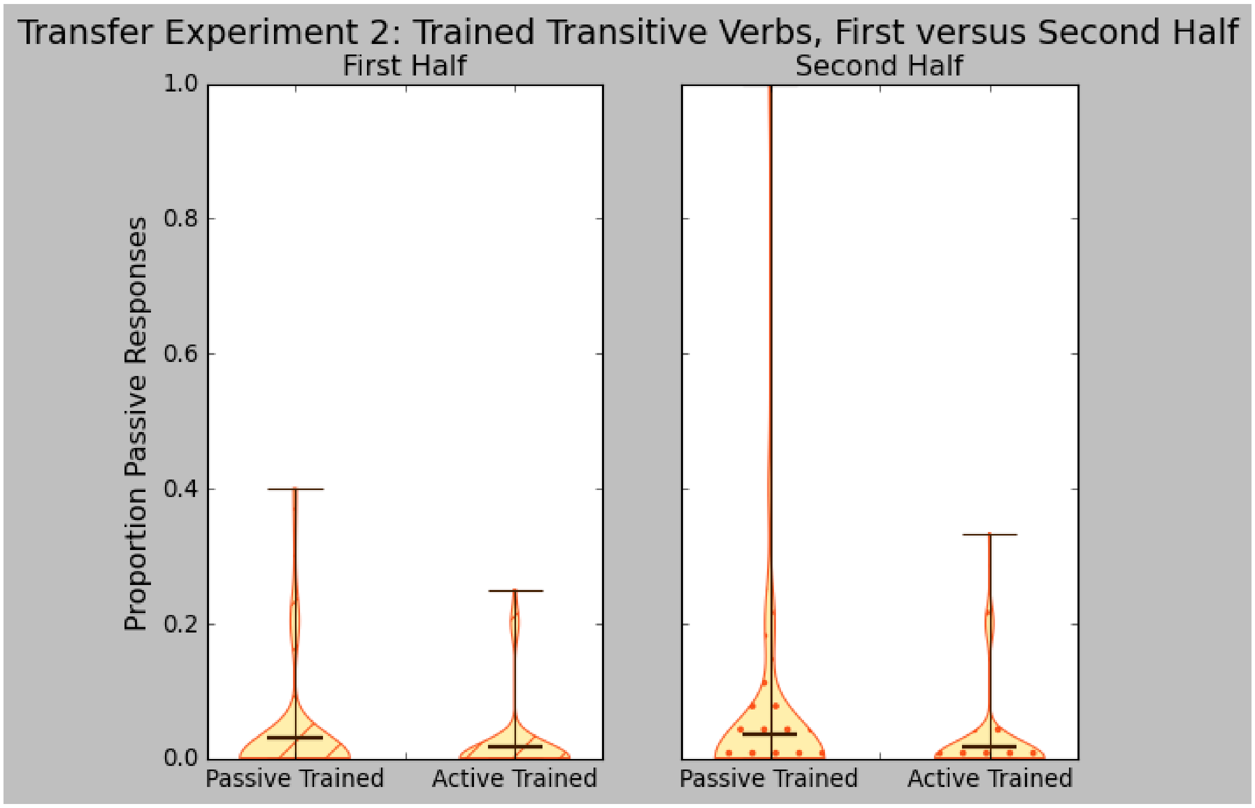

Figure 21: Training effect for untrained transitive verbs, split by half of block.

4.3.3 Conclusions

The purpose of this study was to replicate the first transfer experiment with new dative and transitive verbs. This replication did demonstrate that the trained dative verbs acquired new biases; however, neither the untrained dative verbs nor the transitive conditions showed any effect. In other words, the effect of transfer found for the datives and the effect of training found for the transitives did not replicate. There are several possible interpretations for this finding. It is possible that the transfer effect found in the first experiment was spurious, or that both of these experiments reflect movement around a small true effect size. However, because the second experiment also used different verbs than the first, the more interesting possibility is that these two experiments represent different cases within a single unified system.

For the dative verbs, it is possible that the average amount of semantic similarity dropped between the first and second experiment. The verbs chosen for the first experiment were ship-mail and throw-toss – verbs that most native English speakers would agree are synonyms even without norming data. The verbs selected for the second experiment were hand-pass and take-bring. Intuitively, these verbs seem less similar to each other than the verbs from the first experiment, and this is supported by a numerical difference in the norming data. On a scale from one to five, the average similarity of the dative verb pairs in the first experiment is 4.30, while in the second experiment it is 3.73. Consequently, it is possible that the difference between the two experiments is the result of a graded difference in effect size as the similarity of the two verbs decreases. It is also worth noting that throw-toss and ship-mail are “better” dative verbs than take-bring. “Take,” and to some extent “bring,” both had a strong tendency to be produced in structures other than the dative in this experiment. Besides having more missing data than the other verb pairs, it is also possible that being used in a wider variety of structural contexts changes how learning transfers.

For the transitive verbs, the story is somewhat different. In previous experiments with smaller sample size, there was no evidence of verb bias learning for transitive verbs (Kelley, 2019). The first transfer experiment obtained a numerical difference in the correct direction to indicate a training effect, but it did not reach significance. One explanation for this effect is that transitive verbs cannot be trained to have new verb biases, or that the active/passive alternation does not represent the same kind of alternation as the dative alternation. However, the type of verb used in the first and second transfer experiments offers a second layer of complication. In the first transfer experiment, the passive verbs were theme-experiencer verbs, which are more likely to appear in the passive (Ferreira, 1994). In the second transfer experiment, the verbs were more concrete and were not theme-experiencer verbs, and were extremely unlikely to appear in the passive. Taken together, this suggests a second explanation. It seems that the effect size of learning a transitive verb bias is almost certainly smaller than learning a new dative verb bias. Additionally, the passive is a rare enough structure that production is almost at floor for the second transfer experiment. Consequently, it is possible that this effect is only detectable under the best possible conditions, both when the passive structure is relatively likely to be produced anyway and the sample size of the study is relatively large.

Finally, this study does not replicate the finding from Chapter 3 that the effect of training is larger for dative verbs in the first half of a testing block. It is worth noting that the overall difference is smaller in the replication experiment, and that a considerable number of participants were dropped in both the trained and untrained dative conditions. This is likely due to the tendency of “take” and “bring” to be produced in structures other than the dative, which results in more missing data in the replication. It is also worth noting that the difference between the halves of the block was very small when it was detected in Chapter 3. Consequently, this question is probably best-addressed with a different set of verbs, and carefully considering whether an experiment has the power to detect this effect.

Ultimately, the transfer replication has not settled the issue of whether verb bias learning consistently transfers between semantically related verbs. However, it has brought up a number of new questions. Further work is needed to fully integrate these two data sets in order to better understand if the known differences between these two experiments can explain the differences in their results.

CHAPTER 5: MODEL OF DATIVE VERB BIAS TRANSFER EFFECTS

Chapters 3 and 4 suggest that the effects of transfer of verb bias learning from one verb to another are moderated by a number of factors. These include both class-wide statistics that could potentially explain the difference between the dative and transitive verbs, and differences in semantic similarity and verb bias that might explain why between the dative verbs in Chapter 3 showed transfer and those in Chapter 4 did not. Separating these two factors experimentally is difficult because of the limited number of synonymous alternating dative verbs. Additionally, since it appears that there is no transfer in the passive, it is unclear if using another verb class or alternation would show the same kinds of transfer effects. However, it is possible to dissociate these two factors in modeling, where both the semantic similarity of inputs and the biases of verbs can be determined specifically.

The cognitive model presented in Chapter 2 has the potential to systematically analyze the differences between the studies in Chapter 3 and Chapter 4. Unlike real verbs, inputs to the model can be systematically varied in terms of their overall semantic similarity. Additionally, the biases that the model acquires through "life experience" for these verbs can be definitively known. Because the learning environment is more highly-controlled, the model can begin to answer questions about whether semantic similarity, distributional learning, or some combination of factors is the best explanation for the results in Chapters 3 and 4.

This chapter presents an updated version of the model used in Chapter 2. The model remains a simple feed-forward neural net that learns to assign a verb to either a prepositional dative structure or a double object dative structure. However, the input to the model has been updated to include semantic information. Using this larger and more detailed verb representation, the following studies examine whether the model can account for the differences in the verb bias learning found between Chapter 3 and Chapter 4.

5.1 METHODS

Like the model described in Chapter 2, the model was a three-layer feed-forward model that was constructed using PyTorch (Paszke et al., 2019). Learning was done via backpropagation, with a learning rate of 0.05 and no momentum. This model used hyperbolic tangent functions for its hidden units, and a cross-entropy loss function to minimize loss. As with the model in Chapter 2, this architecture is typically used for categorizing data into two or more categories (Torch Contributors, 2019). In this case, the two dative structures were the two categories. Individual verb representations were either tagged as Category 0 (prepositional dative) or Category 1 (double object dative), and the model compares this tag to the category it predicted for that verb. Activations at the output units were again evaluated using a softmax function, which converts the activation of a unit into the probability that unit's category would be selected (Duda, Hart, and Stork, 2001). As in Chapter 2, graphs present values from the Category1/DO node, which corresponds with the DO bias of the model for that verb representation.

Two primary differences exist between the model described here and the model described in Chapter 2. First, while the model in Chapter 2 received one hot vectors as verb representations, this model received more detailed vectors of semantic features. These vectors were handmade using a combination of semantic features drawn from Levin (1992), and some that were necessary to distinguish very similar verbs from each other (e.g., "throw" and "toss"). The full chart of semantic features and values for each verb can be seen below in Table 6.

Second, the model in Chapter 2 received a vector with three values as input. This model receives a vector with 10 values in order to accommodate the semantic features.

Table 6: Semantic Representations for Transfer Modeling

Verb	Hand Related	Caused Motion	Caused Possession	Sending	Deictic	Ballistic	Path-focused	Force	Recipient-focused	Specificity
Pass	1	0	1	0	0	0	1	0	0	0
Hand	1	0	1	0	0	0	0	0	1	0
Toss	1	1	1	0	0	1	1	0	0	0
Throw	1	1	1	0	0	1	1	1	0	0
Mail	0	1	1	1	0	0	1	0	0	1
Ship	0	1	1	1	0	0	1	0	0	0
Take	1	1	1	0	1	0	0	0	0	0
Bring	1	1	1	0	1	0	0	0	1	0

Training the model involved two steps. The first was the "life experience" phase, where the model acquired verb biases for known verbs. Each life experience phase included 200 epochs of training, during which each verb was seen ten times. For both experiments, the model experienced all eight verbs listed below. The biases below represent how many times each verb was presented with each structure out of the ten trials in each epoch. For example, "toss" was seen nine times in the prepositional dative, and only once in the double object dative.

Table 7: Life Experience Verb Biases

Verb	Percent PD Bias	Percent DO Bias
Throw	100%	0%
Toss	90%	10%
Ship	90%	10%
Mail	90%	10%
Pass	90%	10%
Hand	70%	30%
Take	90%	10%
Bring	60%	40%

The second phase was the "experimental phase". During the experimental phase, two of the verbs received one epoch of further training – one was biased toward the prepositional dative, and the other was biased toward the double object dative. The trained verbs were fully counterbalanced, and the different lists can be seen in the two tables below. Each condition was run 10 times, so that there are a total of 80 individual versions of the model that first received life experience, and then received one condition of experimental experience. Finally, the bias toward each structure of both the trained and untrained verbs was evaluated once at the end of the experimental phase. The bias toward the double object dative was found by evaluating the activation of the appropriate node and using a softmax function to convert this activation to a probability.

Table 8: Study 1 Lists

List	PD Verb	DO Verb
1	Throw	Mail
2	Throw	Ship
3	Mail	Throw
4	Ship	Throw
5	Toss	Mail
6	Toss	Ship
7	Ship	Toss
8	Mail	Toss

Table 9: Study 2 Lists

List	PD Verb	DO Verb
1	Pass	Take
2	Pass	Bring
3	Take	Pass
4	Bring	Pass
5	Hand	Take
6	Hand	Bring
7	Take	Hand
8	Bring	Hand

5.2 RESULTS

The results section below begins by reporting a version of the model that is analogous to the version in Chapter 2. This version of the model has 10 input nodes to accommodate the larger semantic vectors, and 10 hidden units. For both studies, this version of the model strongly predicted a transfer effect. For Study 1, a paired t-test showed that trained verbs were much more biased to occur in the double object dative if they were trained to appear in that structure (Average DO Trained DO Bias: 0.748, PD Trained: 0.183, $t(79)=25.08$, $p<0.05$). Similarly, the untrained synonymous verbs were also more biased toward the structure that their synonym had been trained in, although the effect is somewhat smaller (Average DO Untrained DO bias: 0.532, PD Untrained: 0.267, $t(79)=7.68$, $p<0.05$). Study 2 also found that verbs that were trained in the DO were more biased toward it (DO Trained: 0.569, PD Trained: 0.186, $t(79)=11.64$, $p<0.05$), and that this learning transferred to untrained synonyms (DO Untrained: 0.454, PD Untrained: 0.336, $t(79)=3.84$, $p<0.05$). Graphs for each of these studies are presented below. One important note is that the larger training effect in Study 1 also results in a stronger transfer effect than in Study 2. Second, since the life experience is equivalent between these two models, this must result in differences in surprisal between the verbs used in Study 1 and Study 2. Consequently, one potential explanation for the transfer differences between the experiments in Chapter 3 and Chapter 4 is the amount of surprisal created by individual verbs during training.

While the model replicates aspects of the behavioral data, its effect sizes are considerably larger than those found in the behavioral data, and the model predicts transfer in both experiments. Meanwhile, the behavioral data suggests that there should only be transfer in first study and not the second. There are two potential adjustments to the model that could bring its performance more in line with the behavioral data. First, the representations of the verbs themselves could be adjusted. Chapter 4 demonstrated that there is generally a lower degree of similarity between the verbs in the replication transfer study as opposed to the original study. Consequently, a smaller degree of transfer could also be interpreted as the result of a lower degree of similarity between the verbs. Second, the number of hidden units could be decreased. More hidden units allow for a more complex category boundary, while a model with fewer hidden units must create a simpler one (Duda, Hart, and Stork, 2001). In order to learn effectively with this simpler category boundary, the model may have to create representations of the verbs that are more different from each other. The greater differences between these representations would result in less transfer from trained verbs to untrained ones. The first model study presented below implements a version of the model with less similar verb classes, and the second demonstrates how the model behaves when it includes fewer hidden units.

First, results from a version of the model with verb representations that more closely align with the second study are presented. In this version of the study, the verbs "take" and "bring," which had the lowest similarity ratings of the four pairs of dative verbs, had their verb representations adjusted to be different in four features rather than one. This is intended to more closely align with the similarities reported in Chapter 4. The original and adjusted semantic vectors are reported below, as well as the Euclidean distance between them.

Table 10: Semantic Features Adjusted for Lower Similarity

Verb	Hand Related	Caused Motion	Caused Possession	Sending	Deictic	Ballistic	Path-focused	Force	Recipient-focused	Specificity
Pass	1	0	1	0	0	0	1	0	0	0
Hand	1	0	1	0	0	0	0	0	1	0
Take	1	1	0	0	1	0	0	1	0	0
Bring	1	1	1	1	1	0	0	0	1	0

Table 11: Semantic Similarity of Verbs in Study 2

Verb Pair	Euclidean Distance	Similarity Judgment
Pass/Hand	1.41	4.26
Take/Bring	2.0	3.19

Overall, it does not appear that reducing the similarity of the semantic vectors alone can account for the differences between the first behavioral transfer experiment and the replication. Using these semantic features in a model with 10 hidden units, the model finds both a large training effect (DO Trained: 0.630, PD Trained: 0.207, $t(79)=15.46$, $p<0.05$) and large effect of transfer to the untrained synonyms (DO Untrained: 0.487, PD Untrained: 0.320, $t(79)=5.36$, $p<0.05$). The differences in the means and the value of the t-statistic are both comparable to the first version of the model. Reducing the number of hidden units together with reducing the similarity of the transfer verbs, however, does result in a model that more closely resembles the behavioral results in Chapter 4. This model finds an effect of training (DO Trained: 0.462, PD Trained: 0.279, $t(79)=7.39$, $p<0.05$), but the effect of transfer is much reduced (DO Untrained: 0.395, PD Untrained: 0.347, $t(79)=2.04$, $p<0.05$).

Second, results from the version of the model with fewer hidden units are presented. This model is identical to the model described in the methods section, except that it has three hidden units rather than ten. This model finds a greater amount of DO bias for verbs that are trained to prefer that structure (DO Trained: 0.462, PD Trained: 0.226, $t(79)=9.57$, $p<0.05$), as well as a smaller amount of transfer from trained verbs to their untrained synonyms (DO Untrained: 0.364, PD Untrained: 0.238, $t(79)=4.55$, $p<0.05$). For the second study, this model finds the same training effect as previous experiments (DO Trained: 0.467, PD Trained: 0.261, $t(79)=6.50$, $p<0.05$), but does not find an effect of transfer from the trained verbs to the untrained verbs (DO Untrained: 0.392, PD Untrained: 0.377, $t(79)=0.535$, $p>0.05$). While these effect sizes are still larger than the behavioral data, the pattern of changes in effect sizes is much closer to the behavioral results in Chapters 3 and 4.

Since the first and second transfer studies are otherwise identical, this difference must arise from differences in the life experience phases of the training. There are two possible reasons why this could be happening. Verbs in the first study were generally strongly biased toward the prepositional dative, and consequently experienced a significant amount of surprisal when they were trained toward the double object dative. By contrast, the verbs in the second study were more equibiased, and therefore experienced less surprisal during training. The additional separation provided by the smaller number of hidden units also suppresses the amount of transfer from the trained to the untrained verbs. In the second study, where there is less learning in the first place, this suppressed transfer results in no transfer from the trained to the untrained verbs.

However, it is also the case that the verbs in the second experiment had more different biases than those in the first experiment. In first transfer study, all verbs were approximately equally biased toward the prepositional dative. In the second transfer study, some verbs were more equibiased, and others were more PD-biased. The need to learn these more unique verb biases could have led to a greater need to separate these representations at the level of the hidden units, which consequently would also prevent transfer. It is certainly possible that both the representations of the verbs and surprisal are interacting to create the differences in the modeling of the first and second transfer experiments. Further work is needed to separate these two explanations.

Based on these results, it appears that reducing the number of hidden units results in the model that most closely resembles the behavioral results described in Chapters 3 and 4. Only reducing the similarity between the verb representations cannot account for the pattern of behavioral results; the transfer effect is essentially the same size. While a model that has both a reduced number of hidden units and representations with more realistic similarities also behaves like the human participants in Chapter 4, the primary driver of this is the reduced number of hidden units.

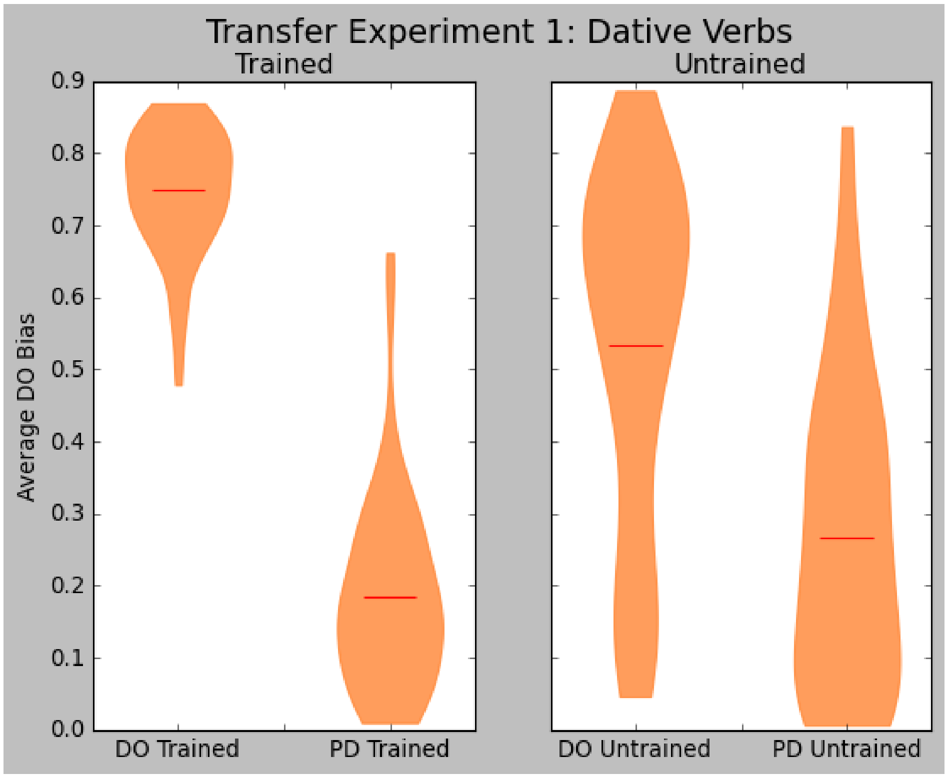

Figure 22: Modeling results for first transfer experiment, using initial semantic vectors and 10 hidden units.

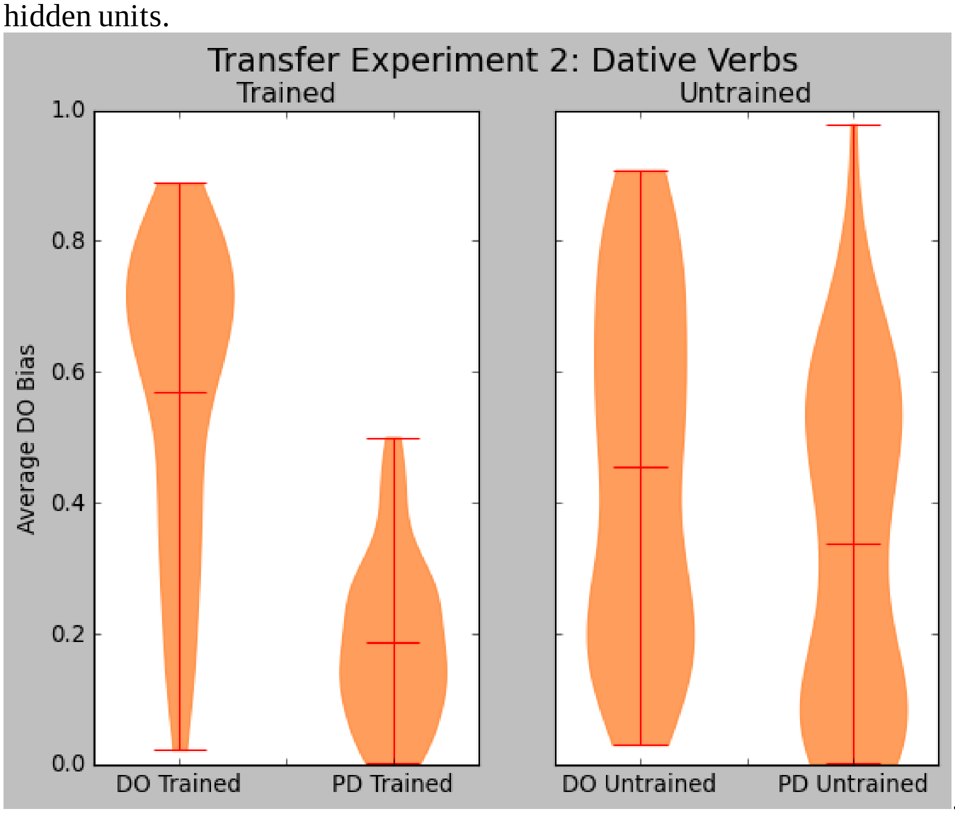

Figure 23: Modeling results for second transfer experiment, using initial semantic vectors and 10 hidden units.

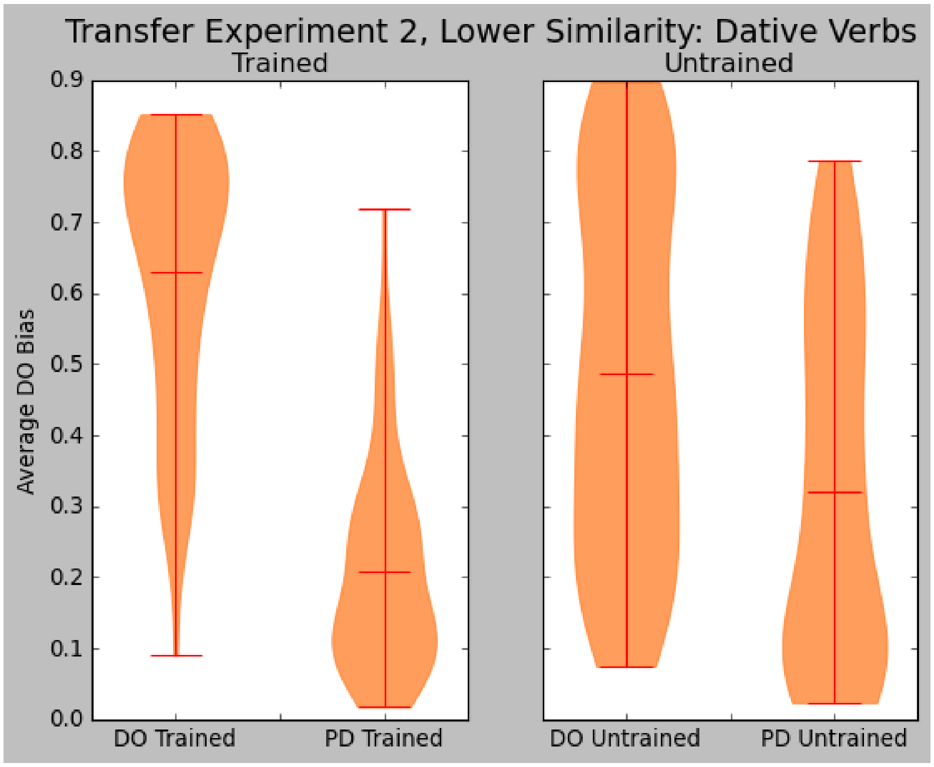

Figure 24: Modeling results for second transfer experiment, using less-similar semantic vectors and 10 hidden units.

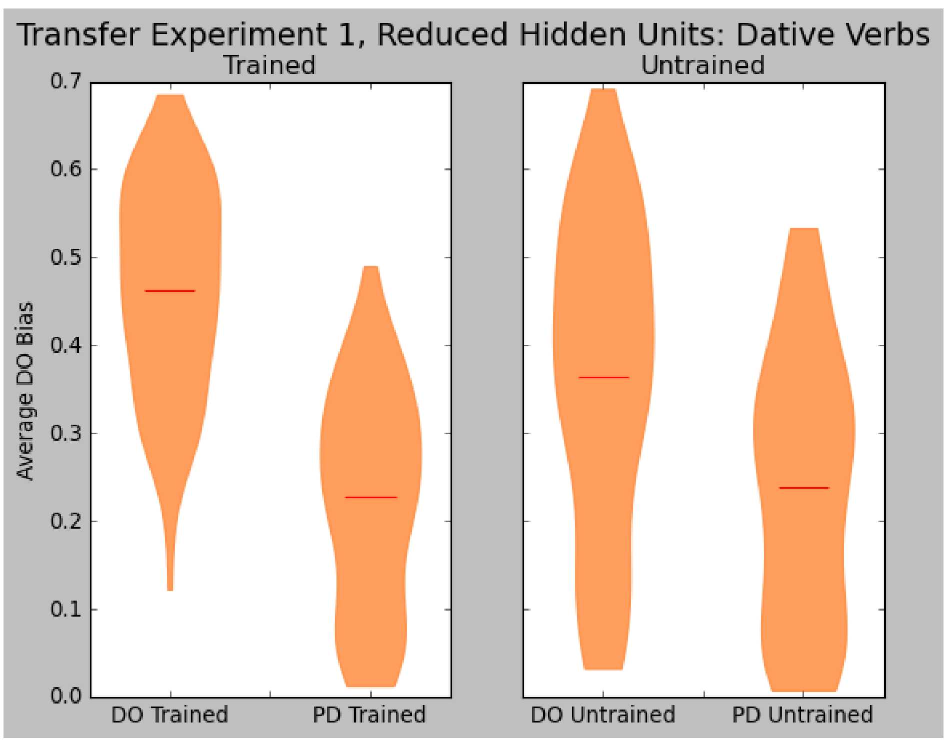

Figure 25: Modeling results for first transfer experiment, using initial semantic vectors and 3 hidden units.

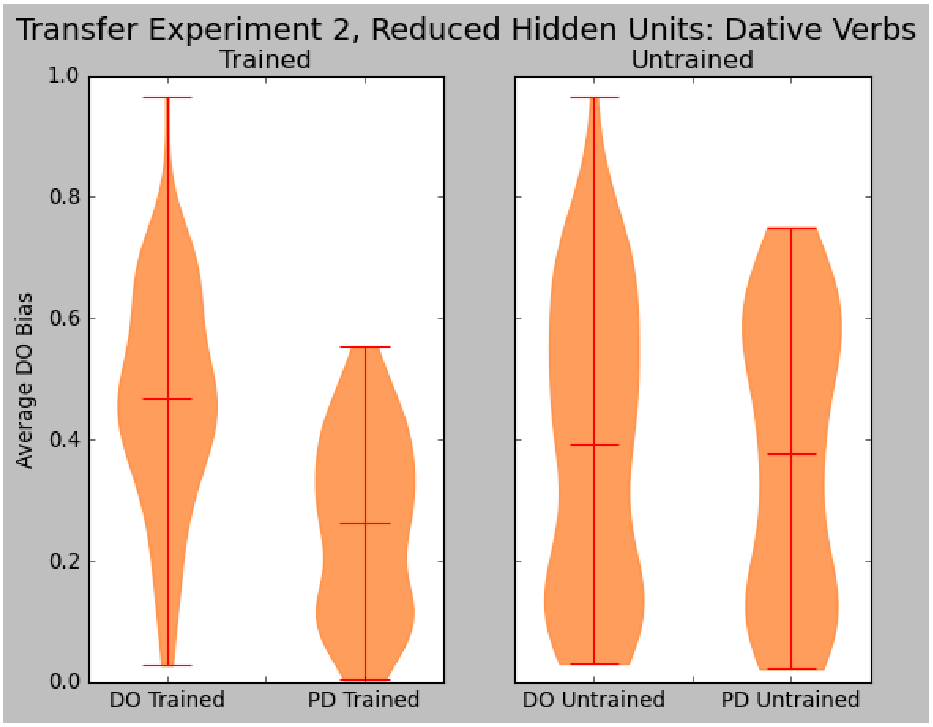

Figure 26: Modeling results for second transfer experiment, using initial semantic vectors and reduced hidden units.

5.3 CONCLUSIONS

With some adjustments, a variant of the model described in Chapter 2 can also account for the behavioral results described in Chapters 3 and 4. Chapters 3 and 4 find that under some conditions, there is a transfer of learning from verbs that are trained to appear in a particular syntactic structure to their untrained synonyms. However, this transfer is not universal, and the effect is somewhat small. A model with the same number of hidden units as input units does not account for these results well, showing both training and transfer in all conditions. However, models with a reduced number of hidden units successfully replicate the behavioral results, showing transfer in some cases and not others. By contrast, varying the amount of similarity between the semantic vectors does not result in a pattern of results that resembles the behavioral results.

This model architecture reveals a number of points about the behavioral data. First, it shows how the overlapping representations of the different verbs allow for transfer. Rather than occurring at the level of the verbs themselves, the crucial similarities in transfer lie at the level of hidden units. This is shown by the finding that reducing the number of hidden units, but not the similarity of the input vectors, is the most important factor in replicating the behavioral data. The main effect of reducing the number of hidden units is that the hidden unit representations of the verbs would become more different from each other, which would be necessary for the model to successfully learn individual biases for each verb. Consequently, this finding suggests that the representations of verbs in the human mind are, in general, not similar enough to each other to allow for this kind of passive transfer of learning. This may be the case when verbs are extremely similar (e.g., throw and toss), but in general even verbs that describe relatively similar actions (e.g., pass and hand) maintain differentiated representations that prevent transfer.

Second, the model suggests that life experience is an important mechanism for explaining more complex findings. The primary difference between the behavioral studies in Chapters 3 and 4 is the biases of the individual verbs. While the verbs in Chapter 3 were strongly biased toward the double object dative, the verbs in Chapter 4 were somewhat more evenly split between the two dative alternatives. Whether the differences between the findings in Chapter 3 and Chapter 4 are driven by a larger amount of surprisal, the more varied biases of the verbs in Study 2, or some combination of the two, it appears that the initial biases of these verbs are an important part of explaining the differences between these experiments.

Additionally, the findings of the transfer model suggest why transfer may not be common between dative verbs. First, it appears that most verbs may not have representations that are similar enough to facilitate transfer in most cases. This is determined by observing that the model with three hidden units rather than ten better models the behavioral data, which will result in greater separation between the representations at the level of the hidden units. However, this also makes sense from a behavioral perspective. Verbs that appear synonymous could appear in different contexts, with different frequencies, and have multiple meanings that do not fully overlap with each other. The verbs “hand” and “pass” from this data set are a good example of this phenomenon. While handing or passing someone an apple could describe the same situation, handing versus passing someone a football could be totally different actions. Presumably, the representations of “pass” and “hand” contain this information, and this could separate the representations enough to prevent transfer. This could also explain why some verbs with similar meanings, like “give” and “donate”, maintain very different structural preferences.

Second, it is likely that the level of surprisal needed to see a transfer from a verb to its synonyms does not occur very often. The biases of verbs are gathered from a speaker’s lifetime of experience hearing verbs in a specific bias. While it is not impossible that a strongly PD-biased verb could be heard in the double object dative multiple times in a row in a natural setting, this is much more likely to happen in an experimental setting. Consequently, while transfer can be found in some experimental settings, it may be a somewhat uncommon occurrence simply because of the combination of factors needed to see a measurable effect.

CHAPTER 6: MODEL OF TRANSITIVE VERB BIAS LEARNING EFFECTS

Generally, the literature supports the idea that dative verbs can rapidly acquire new biases in a laboratory experiment. This finding has been replicated several times with different sets of dative verbs (e.g., Coyle and Kaschak (2008), Qi, Yuan, and Fisher (2011)), suggesting that the underlying effect is a general property of dative verbs. Additionally, similar training effects have been obtained in comprehension using verbs that alternate between instrument (Tickle the frog *with the feather*) and modifier (Tickle *the frog with the feather*) interpretations (Ryskin et al., 2017). Consequently, it is tempting to conclude that many other kinds of common alternations should be able to be updated over the course of a single experiment.

However, transitive verbs that alternate between active and passive structures have remained resistant to this kind of learning. Despite finding that datives could update their verb biases, Kelley (2019) found twice that transitive verbs do not acquire new biases toward passive or active structures. Additionally, this finding was replicated twice in Chapter 4, which found that transitive verbs displayed neither acquisition of new verb biases nor transfer of those verb biases between semantically related verbs. There are multiple possibilities as to why this is the case, which stem from the differences between transitive and dative verbs and their alternations. Experimental evidence suggests two explanations that could account for these experimental findings.

The first potential explanation is that the distribution of transitive verbs prevents the learning of individual verb biases. Evidence that this could explain the behavior of transitive verbs comes from experiments that show that the amount of alternation in a verb class can change what distributions govern production (e.g., Wonnacott et al., 2008; Lin, 2020). For example, learning new verb biases succeeds when familiar English dative verbs are presented with new biases, but does not succeed when the majority of verbs alternate between structures and only a few are biased (Lin, 2020). Because the only difference between these conditions is the distribution of the verbs, it follows that the distribution must discourage the learning of individual biases. However, the mechanism by which this works is less clear. One possibility is that when verbs are a poor cue to structure, the class emerges as a better cue. Consequently, the learner acquires the general tendency of verbs in the class, rather than the preferences of individual verbs. The other possibility is that a distribution where many verbs alternate creates catastrophic interference. As a result, the model fails to acquire the biases of individual verbs, but does learn the overall frequency of the syntactic structures used by the verb class.

The second explanation suggests that cues block each other in a more concrete manner. In this formulation, structures can be predicted by both sentence meanings and individual verbs. In some cases, either the verbs or the meanings are more predictive, in which case learning about one cue seems to block the other. For example, when the structures of English datives are highly predictable based on the verb in the sentence, participants learn to update verb biases (Thothathiri and Braiuca, 2021). However, when Thothathiri and Braiuca instead made the meaning of sentences a better cue to structure, participants instead learned to produce structures based on the events a sentence was describing. Overall, this suggests that there are multiple cues that may allow for contingency learning, and that verb bias learning and updating are blocked if verbs are less predictive than some other cue.

Although the experimental work so far has been done on either English datives or novel verbs, both of these explanations have something to say about the English passive/active alternation as well. Transitive verbs that alternate between the active and passive are distributed differently than dative verbs. First, the vast majority of transitive verbs can undergo this alternation, while the group of alternating dative verbs is more constrained (Cureton, 1979; Levin, 1993). Second, creating a passive involves producing a patient in the subject position, which may be triggered by discourse or meaning that makes the patient more accessible (e.g., Olson and Filby, 1972; Thompson, Ling, Myachykov, Ferreira, and Scheepers, 2013). Consequently, both the distributional hypothesis and the cue-blocking hypothesis have potential to explain the empirical results for the transitives found in Chapter 5.

The following studies aim to see whether experiencing data with these distributional features can be used to replicate the finding that transitive verbs do not acquire new biases. These studies address the two types of verbs used in previous experiments. One type are standard transitive verbs, which rarely alternate to the passive, and theme-experiencer verbs that alternate to the passive more frequently (e.g., Ferreira, 1994). Life experience for these studies will contain two new features. First, transitive verbs will all have the same verb biases, with standard verbs occurring in the passive 2% of the time, and special theme-experiencer verbs occurring in the passive 20% of the time (Ferreira, 1994). Second, transitive passive verbs will have a "focus" feature that perfectly predicts use of the passive, and indicates that the patient of the sentence has been moved to the topic position. Together, these features should result in a smaller or absent effect of training for transitive verbs, and an elimination of the paired training and transfer effects seen in the dative verbs. Finally, these effects should be confined to transitive verbs; dative verbs from the same life experience set should retain the ability to learn new verb biases. In order to address each of these points, three modeling studies were run. The first addresses the transitive results of the first transfer experiment, and the second models the results of the second transfer experiment. Finally, the third confirms that dative verbs in these models retain the ability to acquire new verb biases.

6.1 STUDY 1

6.1.1 Methods

Like the models used in Chapters 2 and 5, results were modeled using a three-layer feed-forward neural net implemented in PyTorch. Unlike the previous models, this one was larger, including fifteen input nodes, four hidden units, and four output units. The input nodes mapped to semantic features that described each "verb" and the "sentence" they were used in. Fourteen were semantic features that described each verb, and one was a feature that indicated whether the patient of the sentence was in the subject position. In other words, the patient-subject unit was 1 for passive sentences, and 0 for all other sentence structures. Each input verb can be seen in Table 12. Hidden units used hyperbolic tangent activation functions, while output units minimized using a cross-entropy loss function. The four output units corresponded to the four possible structures that a verb could appear in: the prepositional dative, the double object dative, the active, and the passive.

This model experienced a life training phase, followed by an experimental phase. In the life training phase, each verb acquired a life experience bias. The biases for individual verbs can be seen in Table 13. Dative verbs acquired individual biases taken from Lin (2020), while transitive verbs acquired uniform biases taken from Ferreira (1994). Transitive verbs either acquired a standard verb bias, where 98% of structures were active, or a theme-experiencer bias where 80% of its structures were active. Additionally, this life experience phase varied the frequency of individual verbs. Dative and theme-experiencer verbs were presented to the model 10 times each, while standard transitive verbs were presented 50 times each. This asymmetry was intended to re-create the fact that standard transitive verbs are very common in English.

In the first study, the experimental phase involved the training of two transitive theme-experiencer verbs, and the subsequent testing of four verbs. For instance, “anger” would be trained to appear in the active, and “captivate” would be trained to appear in the passive. At test, the model’s bias to categorize each of these verbs would be assessed by passing the activation of the output units through a softmax function, which converts these activations to a probability. The bias for two semantically-related but untrained verbs, “enrage” and “intrigue”, would also be Each trained verb was presented ten times, and the bias of both the trained and untrained verbs were assessed once at the end of training.

Finally, in order to demonstrate the effect of life experience on learning, models were exposed to either zero, fifty, or two hundred epochs of life experience. After that, each model received one epoch of the experimental training described above. Figure 27 displays the procedure by which each model was trained and tested in greater detail. All conditions, including the different amounts of life experience and the different experimental lists, were run on different models and are between-subjects. Ninety total models were instantiated for each life experience condition.

Table 12: Input Verb representations

Verb	Dative	Hand-related	Caused Motion	Caused Possession	Sending	Deictic	Ballistic	Path Focused	Force	Recipient-focused	Positive Valence	Negative Valence	Abstract	Intransitive
Pass	1	1	0	1	0	0	0	1	0	0	0	0	0	0
Bring	1	1	1	1	1	1	0	0	0	1	0	0	0	0
Throw	1	1	1	1	0	0	1	1	1	0	0	0	0	0
Give	1	0	0	1	0	0	0	0	0	1	1	0	0	1
Show	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Slice	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Chop	0	1	0	0	0	0	0	0	1	0	0	0	0	1
Push	0	0	1	0	0	0	0	0	0	0	0	1	0	0
Shove	0	0	1	0	0	0	0	0	1	0	0	1	0	0
Anger	0	0	0	0	0	0	0	0	0	0	0	1	1	1
Enrage	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Captivate	0	0	0	0	0	0	0	0	0	0	1	0	1	0
Intrigue	0	0	0	0	0	0	0	0	0	0	1	0	1	1

Table 13: Life Experience Biases

Verb	Percent PD Bias	Percent DO Bias	Percent Active Bias	Percent Passive Bias	Token Frequency
Pass	90	10	0	0	10
Bring	60	40	0	0	10
Throw	100	0	0	0	10
Give	50	50	0	0	10
Anger	0	0	80	20	10
Enrage	0	0	80	20	10
Intrigue	0	0	80	20	10
Captivate	0	0	80	20	10
Slice	0	0	98	2	50
Chop	0	0	98	2	50
Push	0	0	98	2	50
Shove	0	0	98	2	50

Table 14: Study 1 Condition Lists

List Number	Active Verb	Passive Verb
1	Anger	Captivate
2	Anger	Intrigue
3	Enrage	Captivate
4	Enrage	Intrigue
5	Captivate	Anger
6	Captivate	Enrage
7	Intrigue	Anger
8	Intrigue	Enrage

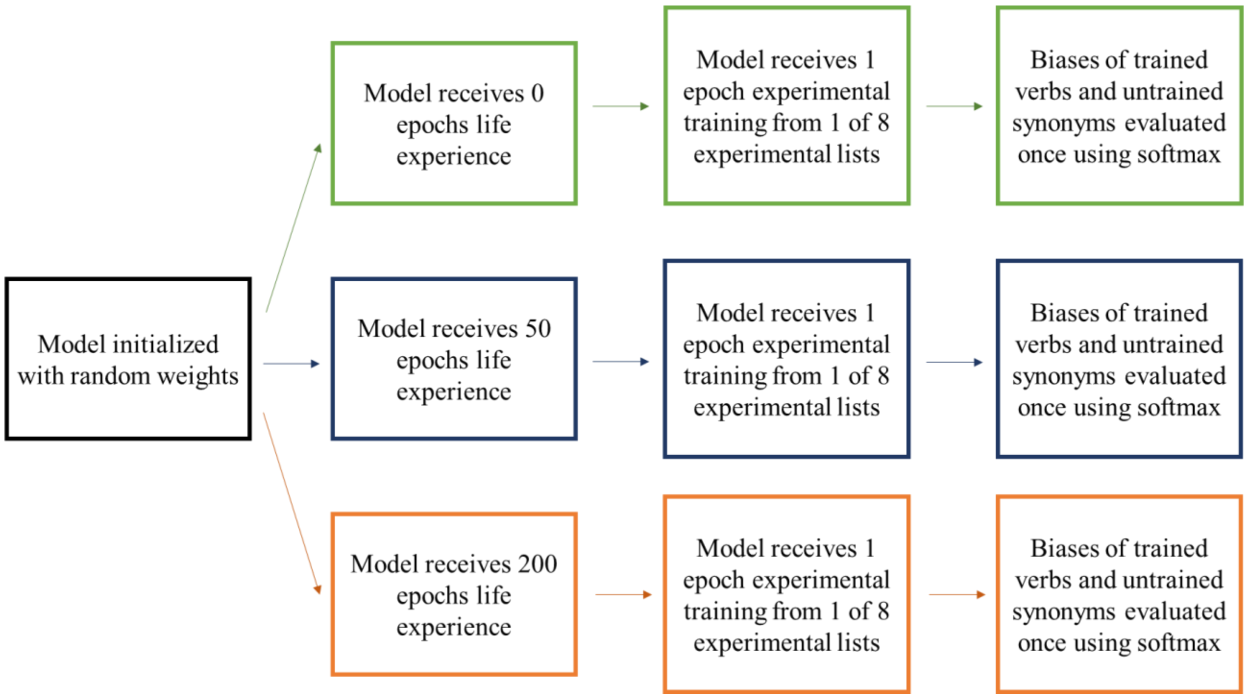

Figure 27: Training regimen for model from Study 1.

6.1.2 Results

Overall, Study 1 demonstrates the importance of life experience to replicate the empirical effects found in the first experiment of Chapter 5. When the model received 0 epochs of life experience and proceeded directly to experimental training, passive-trained verbs are more biased toward the passive than active-trained verbs (Passive Trained Bias: 0.544, Active Trained Bias: 0.258, $t(89)=9.67$, $p<0.05$). This effect also transfers to the untrained semantically-similar verbs, so that verbs with passive-trained synonyms are also more biased toward the passive than verbs with active-trained synonyms (Passive Untrained Bias: 0.484, Active Untrained Bias: 0.323, $t(89)=4.37$, $p<0.05$). The results of this study are visualized in Figure 28 below.

With life experience, the training and transfer effects for the theme-experiencer verbs decline. With 50 epochs of life experience, passive-trained verbs are no longer more passive-biased than the active-trained verbs (Passive Trained Bias: 0.480, Active Trained Bias: 0.470, t(89)=0.18, p>0.05). Additionally, this condition shows no transfer from the trained verbs to the untrained synonyms (Passive Untrained Bias: 0.470, Active Untrained Bias: 0.518, t(89)=0.98, p>0.05). These findings also hold for models which experienced 200 epochs of life experience. These models show neither a training effect (Passive Trained Bias: 0.494, Active Trained Bias: 0.513, t(89)=0.34, p>0.05), nor any transfer (Passive Untrained Bias: 0.500, Active Untrained Bias: 0.501, t(89)=0.06, p>0.05).

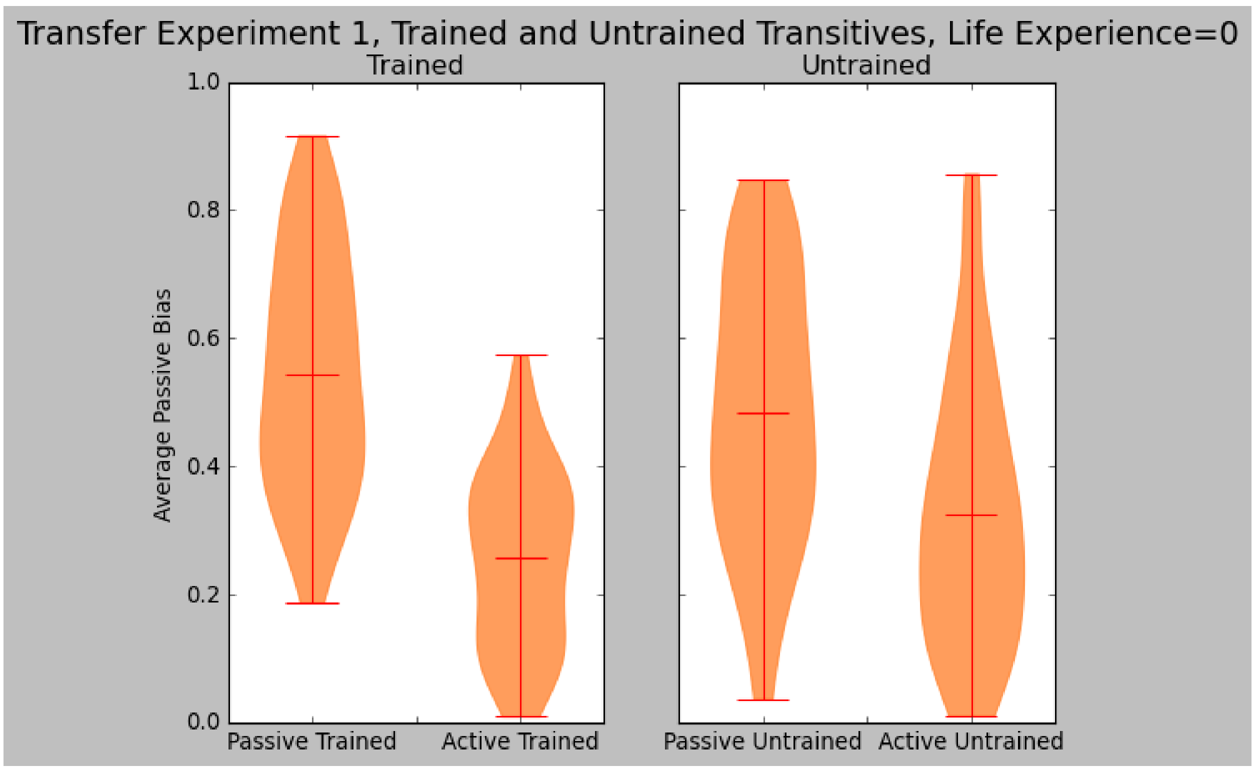

Figure 28: Passive biases of trained and untrained theme-experiencer verbs from models with zero epochs of life experience.

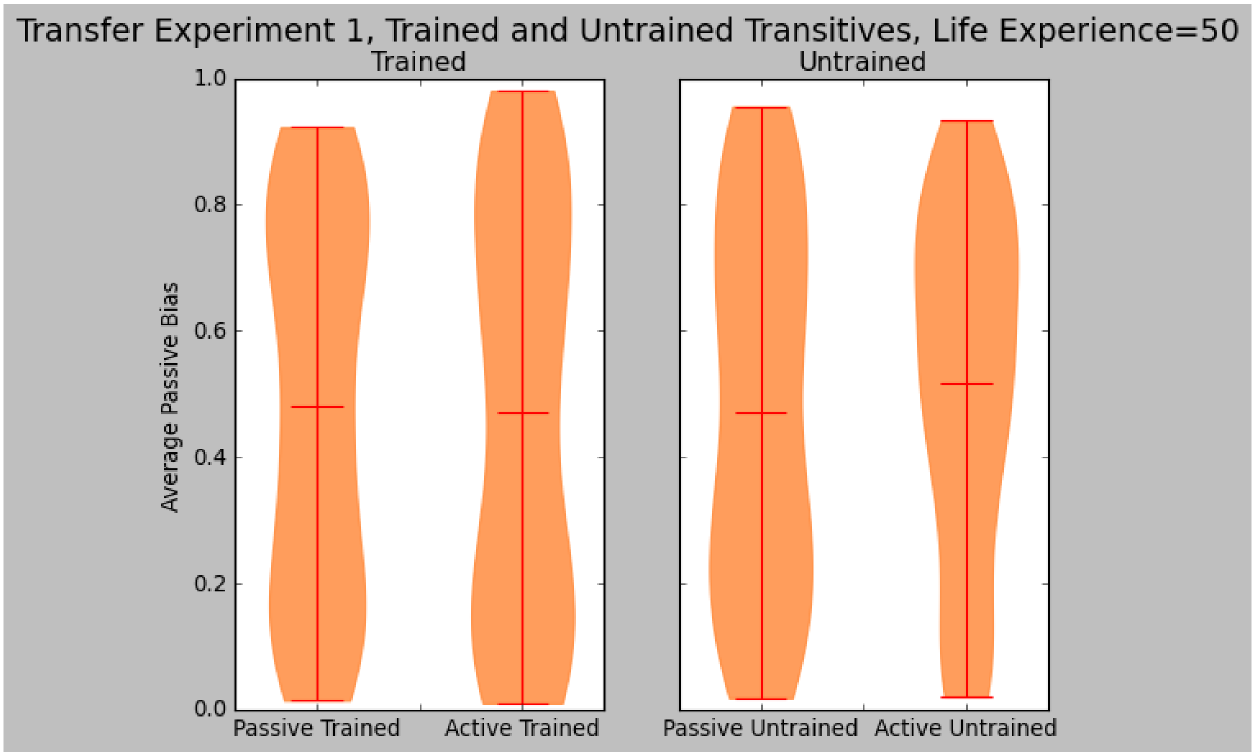

Figure 29: Passive biases of trained and untrained theme-experiencer verbs from models with 50 epochs of life experience.

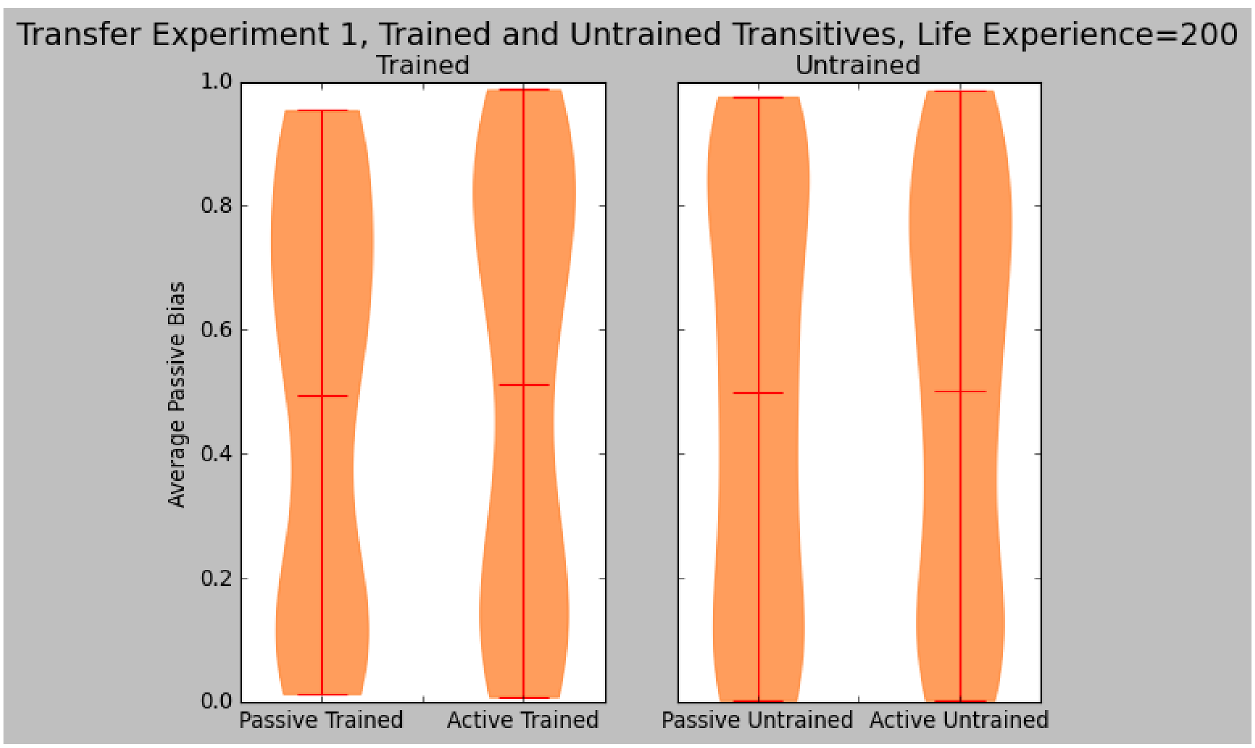

Figure 30: Passive biases of trained and untrained theme-experiencer verbs from models with 200 epochs of life experience.

6.1.3 Conclusions

The empirical results reported in Chapter 4 can be explained by the interaction of two elements. The first is a patient-first feature that perfectly predicts that a particular verb will be categorized into the passive. The second is sufficient life experience with this patient-first feature to recognize that it perfectly predicts the passive. Without the predictive patient-first feature, the transitive verbs would show training and transfer like the dative verbs. Additionally, the 0 Life Experience condition demonstrates that without any life experience, the experimental epoch alone is not long enough to learn that the patient-first feature predicts passive use. Consequently, life experience with a predictive feature is necessary to create the type of effects seen in Chapter 3.

While this feature creates the results presented in Chapter 3, it is also important to verify that it can also replicate the empirical results seen in Chapter 4. Study 2 is intended to show that the predictive patient-first feature is able to replicate not only the results found with theme-experiencer verbs, but also the results found with standard verbs.

6.2 STUDY 2

6.2.1 Methods

The models used in Study 2 are the same as those used in Study 1, and received the same input during the life experience phase. Additionally, all other training and evaluation procedures are also the same, including the comparison between models that experience 0, 50, or 200 epochs of life experience. The primary difference is the verbs that were trained and tested in Study 2.

Instead of training and then testing the theme-experiencer verbs, the experimental epoch used the standard transitive verbs instead. Consequently, each verb was viewed 50 times each during life experience training, rather than 10 times. As in Study 1, counterbalancing resulted in eight experimental lists, which are listed in Table 15 below. For example, if "slice" were trained to occur in the active and "push" were trained to occur in the passive, then these verbs and their untrained synonyms "chop" and "shove" would be evaluated for their passive biases at the end of the experimental training epoch.

Table 15: Study 2 Condition Lists

List	Active Verb	Passive Verb
1	Slice	Push
2	Slice	Shove
3	Chop	Push
4	Chop	Shove
5	Push	Slice
6	Push	Chop
7	Shove	Slice
8	Shove	Chop

6.2.2 Results

Much like Study 1, the results of this study demonstrate that life experience is critical for replicating the effects found in Chapter 5. For the models with 0 epochs of life experience, verbs that are trained in the passive are more passive-biased than the active-trained verbs (Passive Trained Bias: 0.623, Active Trained Bias: 0.282, $t(89)=2.58$, $p<0.05$). The training partially transferred to the untrained synonyms, but the result was not quite significant (Passive Untrained Bias: 0.577, Active Untrained Bias: 0.371, $t(89)=1.92$, $p>0.05$). Results from this experiment are displayed in Figure 31.

With additional life experience, the training effect is no longer present. For models that experience 50 epochs of life experience, there is neither a training effect (Passive Trained Bias: 0.234, Active Trained Bias: 0.231, $t(89)=0.02$, $p>0.05$) nor a transfer effect (Passive Untrained Bias: 0.317, Active Untrained Bias: 0.386, $t(89)=0.43$, $p>0.05$). The absence of training and transfer effects is also the case for the models that experience 200 epochs of life experience (Passive Trained Bias: 0.169, Active Trained Bias: 0.268, $t(89)=-1.82$, $p>0.05$; Passive Untrained Bias: 0.310, Active Untrained Bias: 0.253, $t(89)=0.85$, $p>0.05$). Figures 32 and 33 display the results from the models with 50 epochs of life experience and 200 epochs of life experience, respectively.

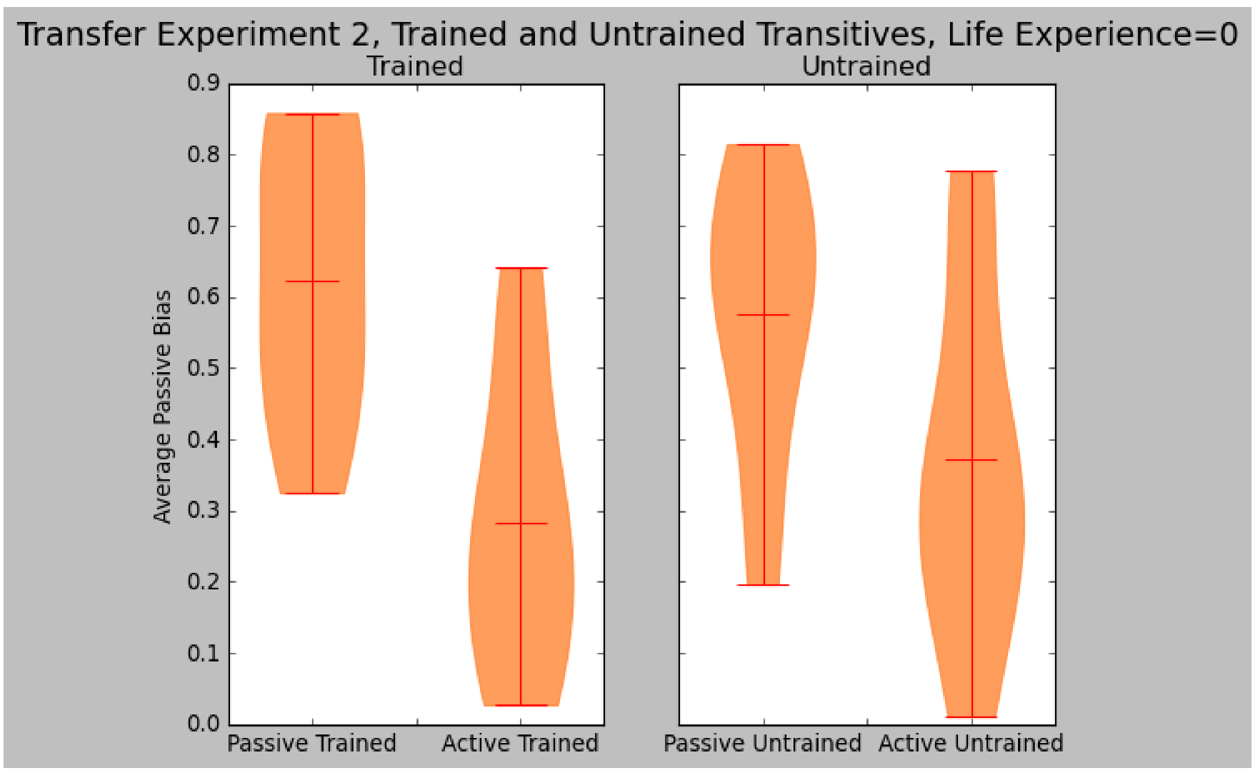

Figure 31: Passive biases of trained and untrained standard transitive verbs from models with 0 epochs of life experience.

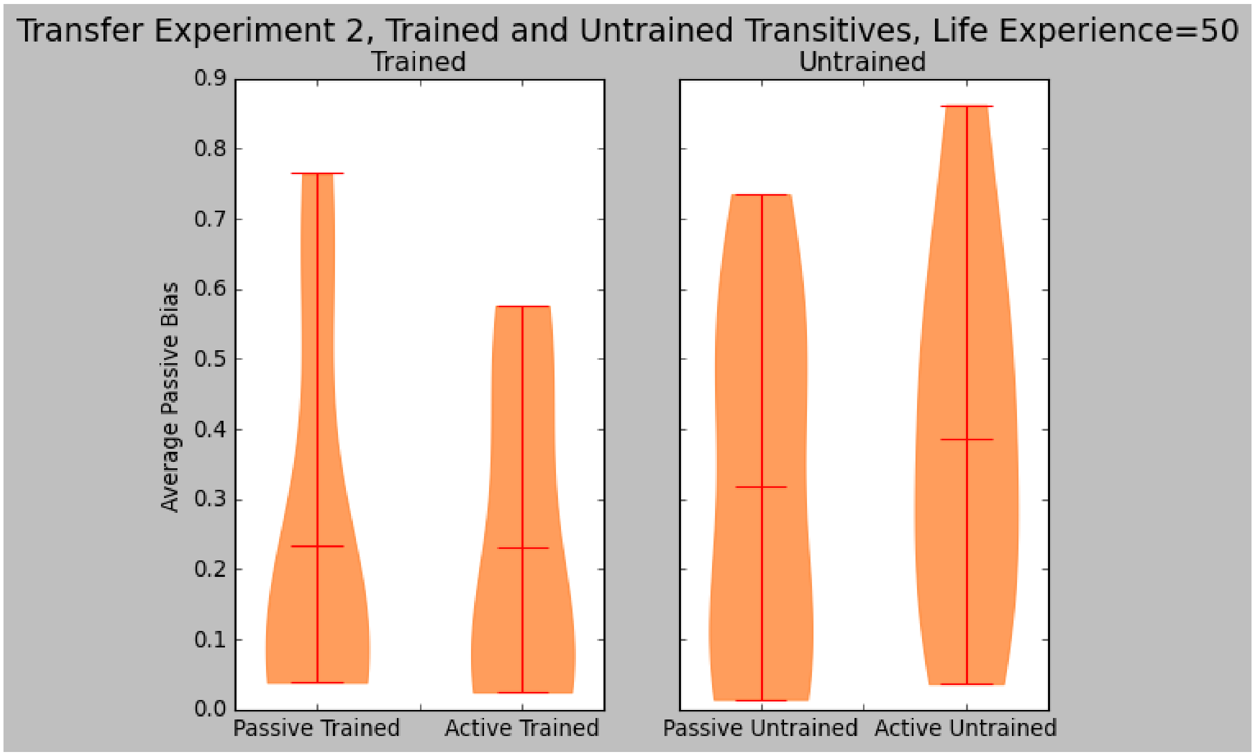

Figure 32: Passive biases of trained and untrained standard transitive verbs from models with 50 epochs of life experience.

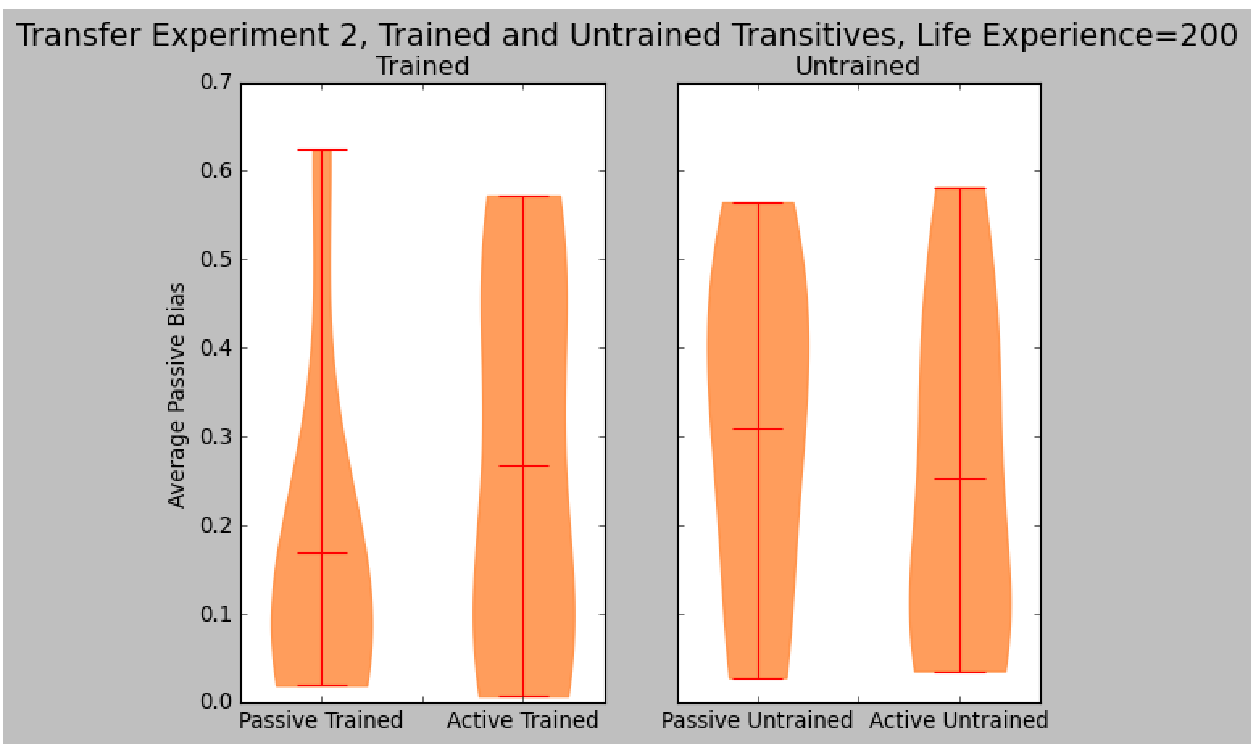

Figure 33: Passive biases of trained and untrained standard transitive verbs from models with 200 epochs of life experience.

6.2.3 Conclusions

Study 2 replicates and extends the results of Study 1. With enough life experience, the model no longer learns transitive verb biases for standard transitive verbs. Again, this seems to be due to the combination of life experience with the transitive verbs and the predictive patient-first cue. One interesting note is that the difference between the untrained verbs in the 0 epochs of life experience condition no longer reaches significance in Study 2. Additionally, the trained verbs in the 0 epochs condition show a reduction in the t-values between Study 1 and Study 2, from 9.67 to 2.58. This is likely due to the slightly sparser representations of the verbs used in Study 2. Life experience allows the model to overcome these initial representational differences, and both models with life experience replicate the empirical results.

One conflict between Study 2 and the empirical results is the difference between the predicted passive bias and the number of passive structures produced by participants. While the model suggests a bias toward categorizing verbs as passive 15% to 30% of the time, passives were produced in Chapter 4 were produced between 1% and 3% of the time. This difference can likely be explained by other production constraints that are not considered by the model. In this context, it is unsurprising that passive production was low in Chapter 4. However, this also explains why the model does not fully replicate these results – it is not sensitive to the presence or absence of discourse that might govern the use of the passive. Although the model is different from the behavioral results, its behavior is explainable based on what is known about passive production.

6.3 STUDY 3

Life experience with the patient-in-subject-position feature is key for preventing the model from learning new active or passive biases for verbs. However, the addition of this feature could also negatively affect learning of all new verb biases. The behavioral results in Chapter 4 indicate that human participants retain the ability to learn new dative verb biases despite their resistance to learning new transitive verb biases. Consequently, this third study is intended to determine whether this model retains the ability to learn new dative verb biases.

6.3.1 Methods

The same model described in Studies 1 and 2 was used, and this model was given 200 epochs of the usual life experience. The primary chance in this study was the type of experimental training, which consisted of learning new biases for the four dative verbs learned during the life experienced phase. For example, “pass” and “bring” might acquire new biases toward the prepositional dative, while “throw” and “give” would acquire new biases toward the double object dative. Each training list is documented in Table 16. Ninety total models were instantiated, so that each list was run fifteen times.

As with the transitive verbs in Studies 1 and 2, the biases for these verbs are evaluated once at the end of the experimental phase. However, since the dative verbs are trained toward either the prepositional dative or the double object dative, their bias toward the double object dative is assessed instead of their bias toward the passive.

Table 16: Study 3 Condition Lists

List	PD Verbs	DO Verbs
1	Pass, Bring	Throw, Give
2	Pass, Throw	Bring, Give
3	Pass, Give	Throw, Bring
4	Bring, Throw	Pass, Give
5	Bring, Give	Pass, Throw
6	Throw, Give	Pass, Bring

6.3.2 Results

The results of Study 3 are presented in Figure 34 below. These findings confirm that the model has retained its ability to acquire new dative verb biases. The verbs that acquired a new DO verb bias were more biased toward the DO than verbs that acquired a new PD bias (DO Trained: 0.618, PD Trained: 0.336, $t(89)=-10.9$, $p<0.05$). Because the life experience of the models in all three studies are equivalent, this finding demonstrates that the model retains the ability to learn dative verb biases even when it no longer learns new active or passive biases.

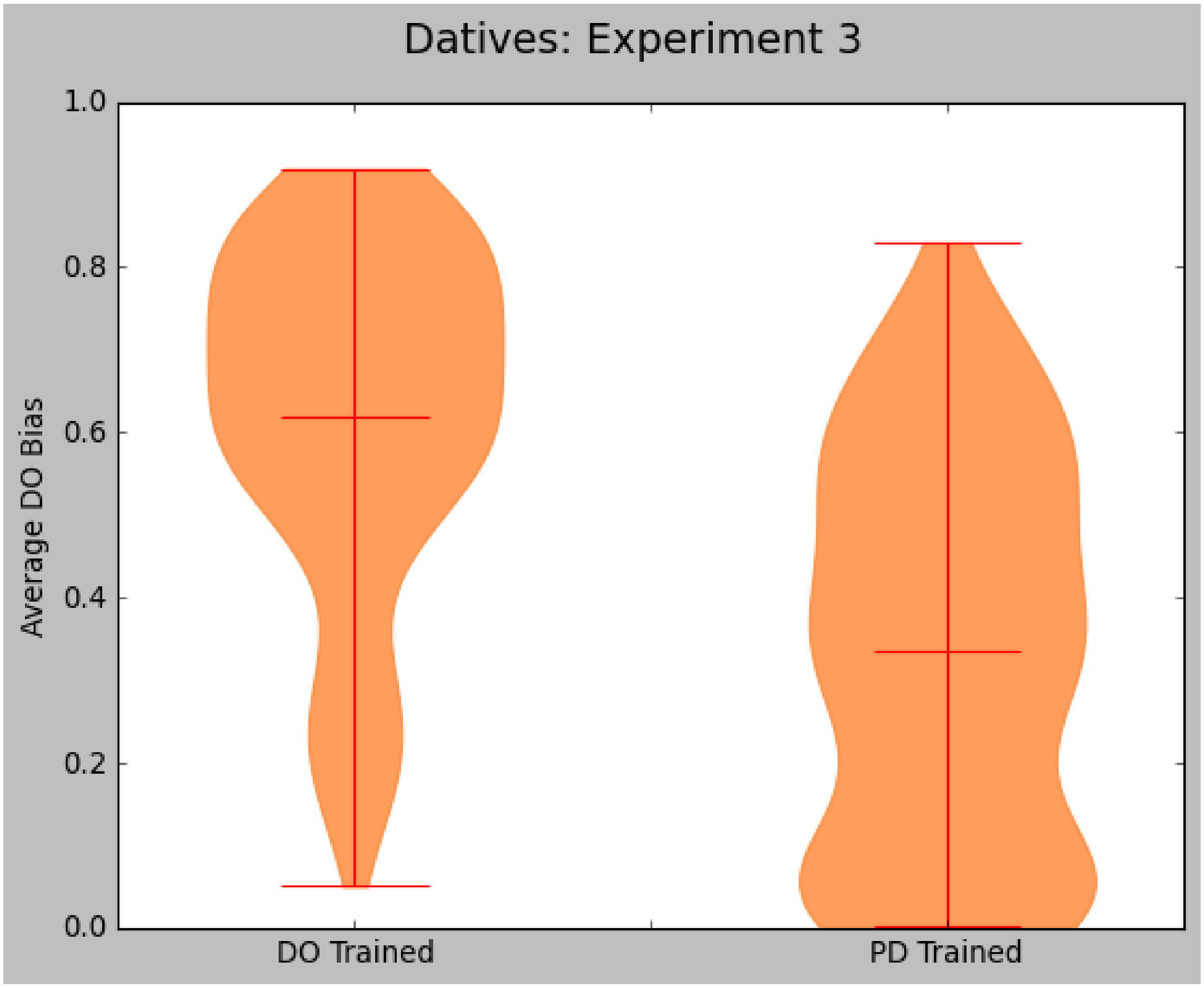

Figure 34: DO bias after training for model with transitive and dative experience.

6.4 GENERAL CONCLUSIONS

Overall, this model demonstrates that a three-layer feed-forward model can account for both sets of empirical results presented in Chapters 3 and 4. While the model in Chapter 5 accounts for the dative findings, the addition of a patient-first feature that predicts the use of the passive allows this similar model to replicate the transitive results from these same experiments. This is true for both the theme-experiencer verbs from Chapter 3, and the standard transitive verbs from Chapter 4. In both cases, life experience with the patient-first feature caused the model not to acquire new verb biases after experimental training. However, this life experience does not prevent the model from learning new dative verbs. Rather, the model has learned that transitive structures are strongly predicted by the patient-first feature, dative structures are predicted by verbs, and that these two classes of verbs are separate.

CHAPTER 7: LEARNING TO CHANGE CUES TO VERB CLASS BEHAVIOR

Verb biases are only one of many cues to the eventual structure of a sentence. Previous work suggests interactions with factors like plausibility in comprehension (e.g., Garnsey et al., 1997), and givenness in production (e.g., Bresnan et al., 2007). Consequently, it is unsurprising that cues to structure compete with one another, and that verb bias is one of the many cues that may eventually win out.

Thothathiri and Braiuca (2021) and Experiment 6 from Lin (2020) were presented in Chapter 6 as two alternative explanations for why verb biases are not learned for transitives. Thothathiri and Braiuca present a situation where learning is blocked by a predictive cue, while Lin (2020) suggests that interference is due to the type of distribution. Chapter 6 suggests that the empirical transitive results are the result of life experience with a cue that predicts the presence of the passive. Specifically, a patient-first cue that perfectly predicts the use of the passive creates model results that resemble the empirical results most closely. When the life experience distribution remains the same, but this cue is not present, the model learns a standard verb bias. Consequently, the life experience distribution alone is not enough to explain the blocking of passive learning in the model, suggesting that the patient-first cue is necessary to block learning in this architecture.

Rather than applying these two possible explanations to other types of verbs, this chapter asks whether the model can successfully replicate the results of Lin (2020) and Thothathiri and Braiuca (2021). The first study replicates Experiment 6 from Lin (2020). In this study, the model is presented with life experience, and then with either a condition where most verbs appear in only one structure (most verbs biased), or a condition where most verbs occur in two structures equally (most verbs alternate). Lin (2020) reports that in the second condition, verb bias learning was blocked. Since there is no cue that is explicitly changed between these two conditions, one possibility is that the most-verbs-alternate distribution does not allow the successful learning of verb biases. If this is the case, then no additional cues should be required to replicate the learning seen in Lin (2020). However, if it is actually the case that participants switch to using a non-verb cue to predict structure when most verbs alternate, then the distributional pattern alone might be able to account for the findings.

The second study replicates two of the three experiments found in Thothathiri and Braiuca (2021). In the first experiment, the majority of verbs occur in only one structure, and only a few verbs alternate equally between the two dative structures used in the experiment. Additionally, the double object dative is always used to describe completed actions, while the prepositional dative is used to describe regular transfer actions. Consequently, both verbs and the event type are highly predictive of the final structure of a sentence. However, in their third experiment, only two verbs occur in only one structure, while the double object dative is still used to describe all completed transfer events. In this case, the event structure is highly predictive of sentence structure while the verb is not, and verb bias learning does not occur. In order to model these experiments, the model will receive life experience with dative verbs, and then will receive input like that in Experiment 1 or Experiment 3. Additionally, a completed-transfer feature will perfectly predict the use of the double object dative in the experimental context. In this case, if the model can shift which cues it uses to predict structures over the course of the experimental phase, then it should be able to replicate the results of both empirical experiments.

7.1 STUDY 1: LIN (2020), EXPERIMENT 6

Lin (2020) Experiment 6 implements a between-groups design that varies whether most verbs are trained to appear in only one structure, or whether most verbs appear equally in two syntactic structures. In the "Most Verbs Biased" condition, participants would see a total of five verbs. Four verbs would appear in only one structure, with two that appeared only in the prepositional dative and two that appeared only in the double object dative. One verb appeared in each structure half the time. In the "Most Verbs Alternate" condition, three verbs alternated between the two structures, one verb appeared only in the prepositional dative, and one verb appeared only in the double object dative.

As seen in Fig. 35, Lin (2020) found that in the Most Verb Biased condition, four- to six-year-olds produced more DO structures in the DO training condition than in the other two conditions. However, this difference was not present in the Most Verbs Alternate condition. Lin takes this as evidence that the distribution of verbs in each condition caused a difference in learning between the two conditions, and suggests that this is evidence that children learn to track structures at the level of verb classes when individual verbs are no longer predictive.

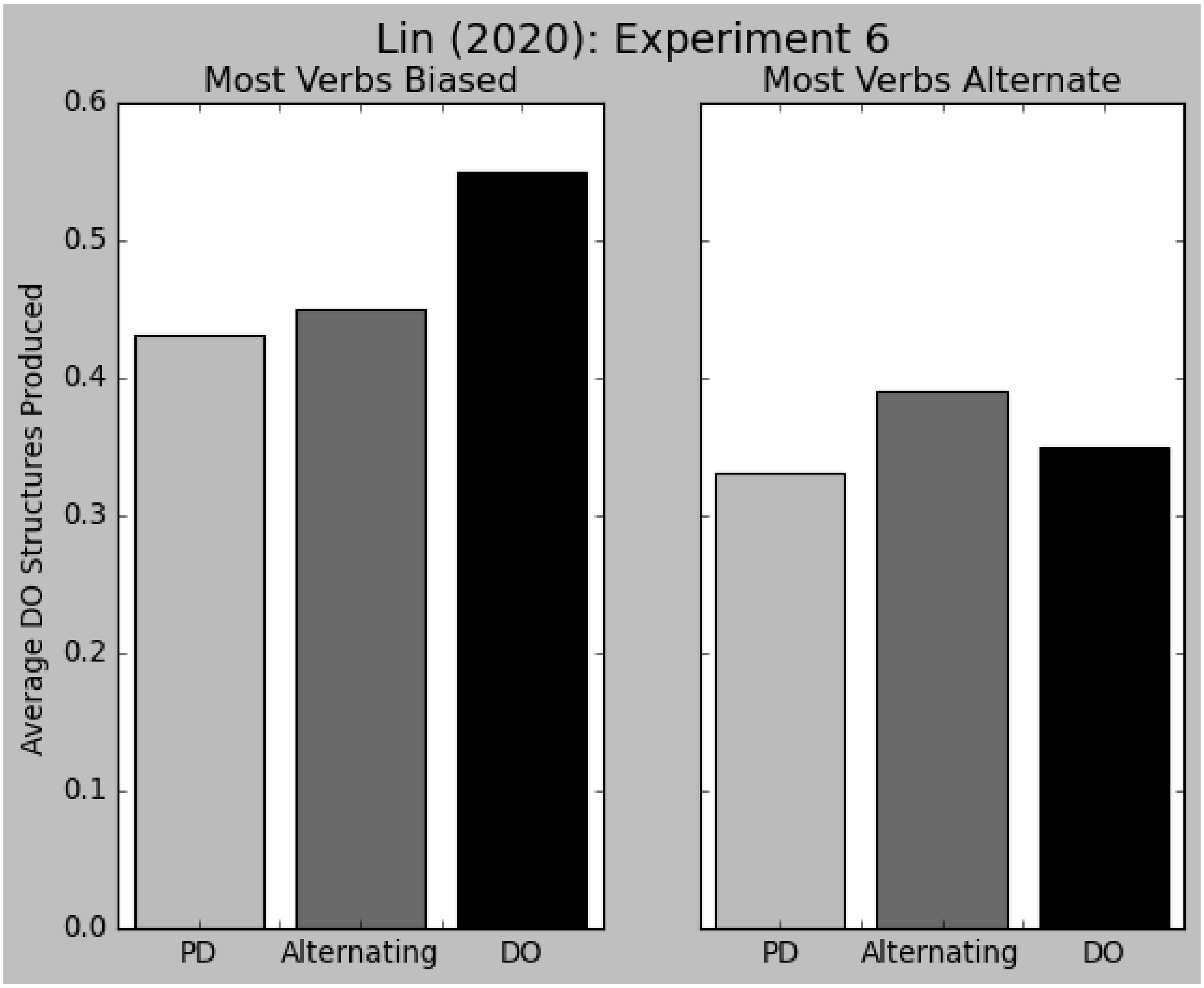

Figure 35: Graph of average double object responses in Most Verbs Biased and Most Verbs Alternate conditions.

In order to model Lin (2020) Experiment 6, the model was first exposed to a life experience phase where it learned biases for each verb. Then, the model would be exposed to training that either replicated the Most Verbs Biased condition, or the Most Verb Alternate condition. To give the model the opportunity to learn about verb classes, the model would receive a distributed verb representation as well as a single feature that indicated if the verb was dative. Finally, the model would be tested by evaluating the bias toward the double object for each verb.

7.1.1 Methods

Like in previous chapters, the model was a three-layer feed-forward neural network. Its hidden layer used hyperbolic tangent activation functions, and the output layer used a cross-entropy loss. The model contained eleven input units, three hidden units, and two output units. The first 10 input units were used to present a distributed representation of a verb to the model, while the last was a feature that was signaled whether a verb was dative or not. Because every verb the model experienced was dative, this feature was always one. This feature was chosen because all of the verbs presented in Lin (2020) were dative. The distributed verb representations used are represented in Table 18 below. The verbs share overlapping semantic features, which may allow the interference needed to obtain the results found in Lin (2020). The output units represented the two possible syntactic alternates, the prepositional dative and the double object dative.

Table 17: Distributed Verb Representations for Lin (2020), Experiment 6

Verb	Hand-related	Caused-motion	Caused possession	Sending	Deictic	Ballistic	Path-focused	Force	Recipient-focused
Pass	1	0	1	0	0	0	1	0	0
Bring	1	1	1	1	1	0	0	0	1
Throw	1	1	1	0	0	1	1	1	0
Give	0	0	1	0	0	0	0	0	1
Show	0	0	0	0	0	0	0	0	1

During life experience, the model learned verb biases for each of the five verbs. Biases were derived from the corpus studies conducted by Lin (2020). Subsequently, the model was either trained using the input from the Most Verbs Biased condition or the Most Verbs Alternate condition found in Lin (2020), Experiment 6. As before, the Most Verbs Biased condition contained one verb that appeared equal numbers of times in both the prepositional dative and the double object dative, while four verbs appeared in only one structure. By contrast, the Most Verbs Alternate condition contained three verbs that appear equal numbers of times in both structures, and only two verbs that appeared in only one structure. As in Lin (2020), both conditions were counterbalanced, and details for each list can be found in Tables 18 and 19 below. Additionally, the instantiation of the models preserved the between-subjects design of Lin (2020). For each condition, fifty individual models were created, given life experience, and then experienced either the Most Verbs Biased condition or the Most Verbs Alternate condition. Training is further explained in Fig. 36 below. After experimental training, the output activations produced by each verb were found, and then evaluated using a softmax function to convert these activations to the probability of that output being chosen (Duda, Hart, and Stork, 2001).

Table 18: Lists for Lin (2020), Experiment 6: Most Verbs Biased

List	DO-trained verbs	Alternating verb	PD-trained verb
1	Bring, throw	Pass	Give, show
2	Throw, give	Bring	Pass, show
3	Give, show	Throw	Pass, bring
4	Pass, show	Give	Bring, throw
5	Pass, bring	Show	Throw, give

Table 19: Lists for Lin (2020), Experiment 6: Most Verbs Alternate

List	DO-trained verb	Alternating verbs	PD-trained verb
1	Pass	Throw, give, show	Bring
2	Bring	Pass, give, show	Throw
3	Throw	Pass, bring, show	Give
4	Give	Pass, bring, throw	Show
5	Show	Bring, throw, give	Pass

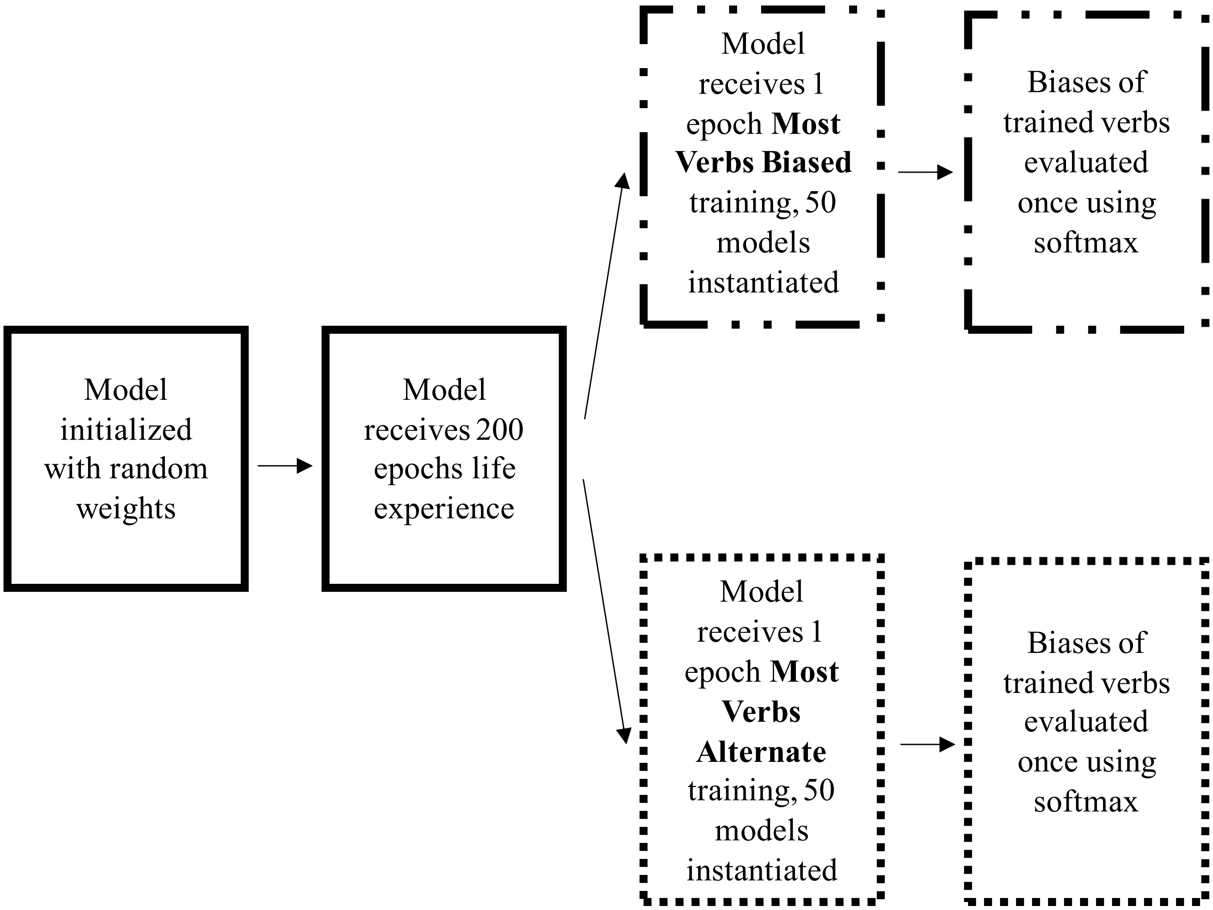

Figure 36: Life experience and experimental procedure for Lin (2020) modeling

7.1.2 Results

Generally, the behavior of the models demonstrated verb bias learning in both cases. For the Most Verbs Biased condition, the average DO bias of the PD-trained verbs was lower than the average bias of the alternating verbs (PD-trained: 0.29, Alternating: 0.45, $t(49)=-4.01$, $p<0.05$). Additionally, the average bias for the alternating verbs was lower than the DO bias of the DO-trained verbs (Alternating: 0.45, DO: 0.59, $t(49)=-2.90$, $p<0.05$). Overall, this suggests that the model has learned that some verbs are more biased to appear in the double object dative than others. This pattern also holds for the biases in the Most Verbs Alternate condition. Once again, the DO bias of the PD-trained verb was lower than the alternating verb, and the DO bias of the alternating verb was lower than the DO-trained verb (PD-trained: 0.30, Alternating: 0.48, $t(49)=-5.05$, $p<0.05$; Alternating: 0.48, DO-trained: 0.61, $t(49)=6.59$, $p<0.05$). Consequently, verb biases were learned in both the Most Verbs Biased and the Most Verbs Alternate conditions, unlike the empirical results found in Lin (2020).

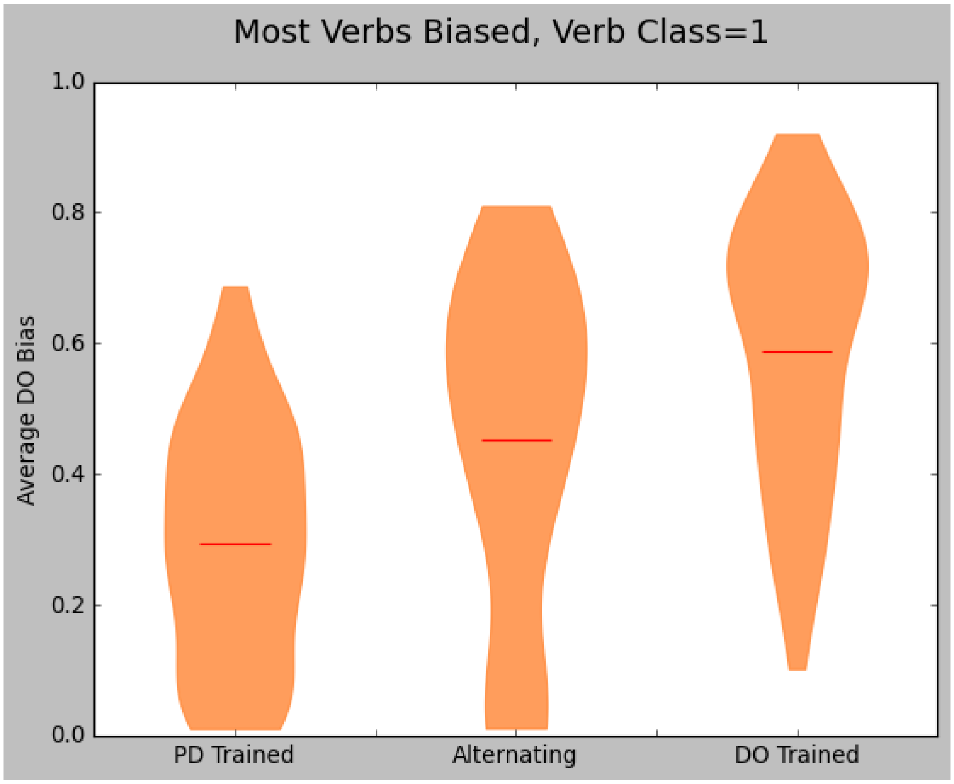

Figure 37: Average DO biases of output units from models exposed to the Most Verbs Biased condition.

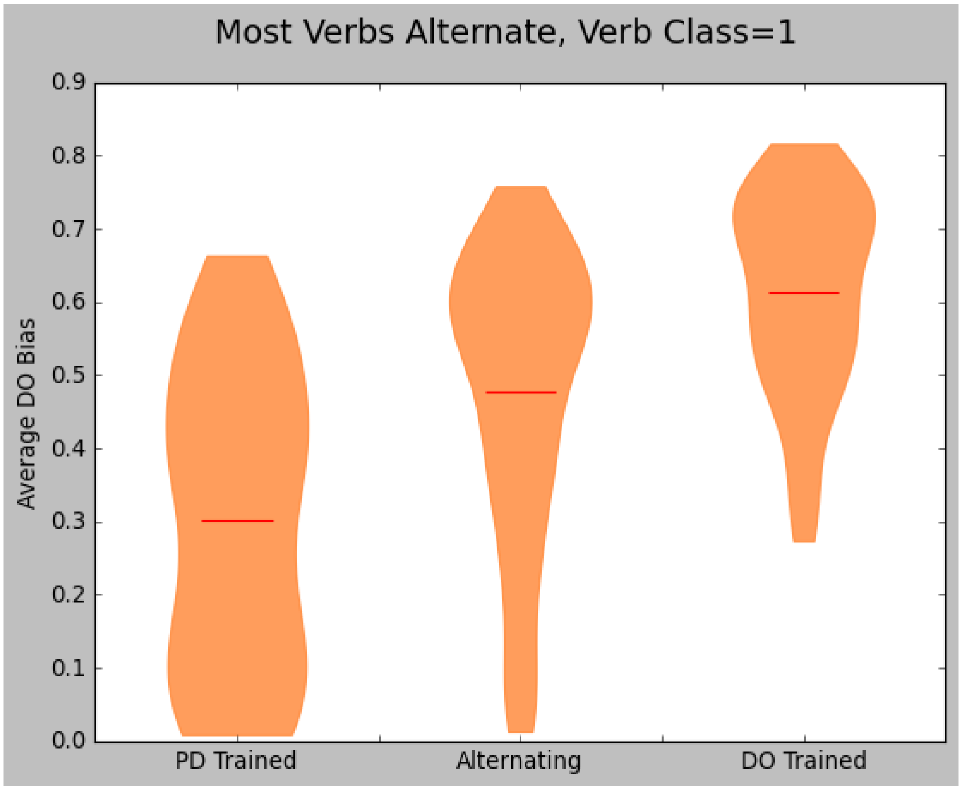

Figure 38: Average DO biases of output units from models exposed to the Most Verbs Alternate condition.

7.1.3 Conclusions

Overall, the model did not succeed in demonstrating differences in learning between the Most Verbs Biased and Most Verbs Alternate condition. While the model successfully learned verb biases in both conditions, the behavioral data from Lin (2020) found that verb bias learning was reduced in the Most Verbs Alternate conditions. This difference suggests something fundamentally different between how the model and the human participants learn about verbs and verb classes.

Lin (2020) suggests that the behavioral results are evidence that participants switched from using verbs to predict structures to using a verb class. While the model technically has all this information, it does not have the ability to learn about whether verbs in general are more or less predictive than other classes of cues. In fact, the input the model receives is not grouped into a “verb” and a “class”, and consequently the model does not have representations that even allow it to consider that some input nodes are of a qualitatively different kind than others. A potential solution to this would be to implement a more complex model architecture that allows the model to both represent these different cues, and to actively shift how much weight each class of cues receives. This could likely be done by adding an attention mechanism to the model, like the one used in Kruschke’s ALCOVE model (e.g., Kruschke, 1990). Future directions for a more complex model and architecture will be returned to in the general discussion, in order to more fully address all relevant modeling and behavioral results.

7.2 STUDY 2, THOTHATHIRI AND BRAIUCA (2021)

Thothathiri and Braiuca (2021) documents three experiments that vary how well verbs and semantics predict structures. Experiments 1 and 3 represent the clearest difference in learning, while the behavior found in Experiment 2 is somewhat ambiguous between these two endpoints and will not be discussed further. In Experiment 1, the majority of verbs occurred in only one structure, either the double object dative or the prepositional dative. A smaller number of verbs appeared in both structures. Additionally, sentences that clearly showed completed transfer actions were always described with the DO, while sentences with potentially incomplete actions were described with the PD. Completeness was determined by watching accompanying video stimuli. In this experiment, completeness was a perfect predictor of structure, but verbs remained a strong predictor as well. Under these conditions, participants updated their verb biases based on what structures the verbs appeared in, which can be seen in Fig. 39.

In Experiment 3, Thothathiri and Braiuca (2021) made verbs a significantly worse predictor of structure, but maintained completed transfer actions as a perfect predictor. In order to do this, they raised the number of verbs that occurred equally in both structures, and decreased the number of verbs that occurred in only one structure. Using this training structure, participants no longer learned verb biases, as can be seen in Fig. 39. Instead, the two verbs that occurred in only one structure were no longer significantly different from each other. Thothathiri and Braiuca explain this as a shift from using verbs to predict structures to using semantics to predict structures.

In order to model these two experiments, each model was first exposed to a life experience phase. During this life experience phase, the model learned both distributed representations of verbs and whether a particular sentence showed a completed action or not. Then, an individual model would be exposed to either the experimental training seen in Thothathiri and Braiuca (2021) Experiment 1, or the training from Experiment 3. Finally, the biases toward the double object dative were evaluated for each model, in order to replicate the dependent variable measured in Thothathiri and Braiuca.

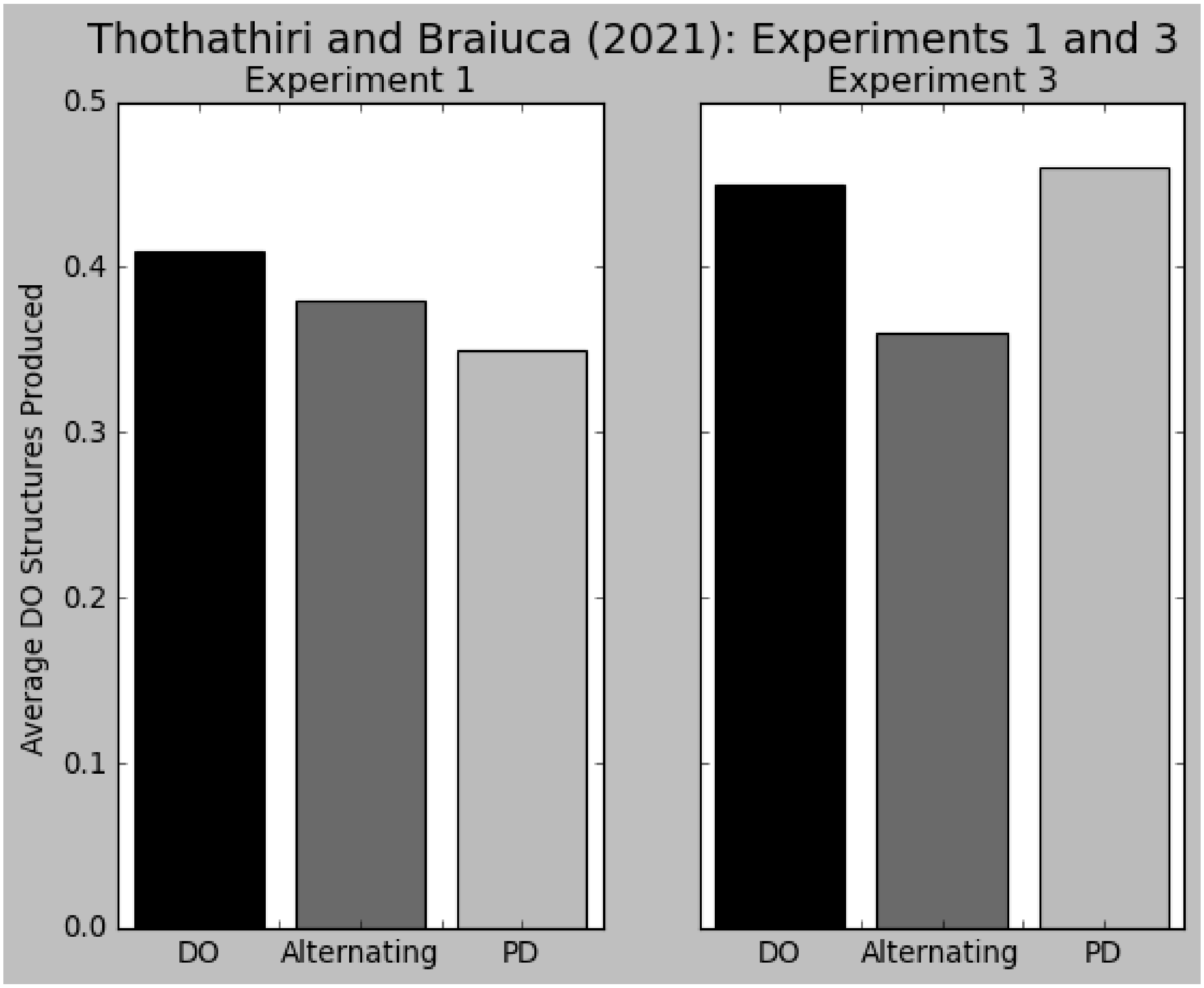

Figure 39: Results from Thothathiri and Braiuca (2021), showing the different numbers of DO structures produced in different training conditions. Experiment 1 demonstrates a standard verb bias effect, while in Experiment 3 this effect is no longer present.

7.2.1 Methods

The model used contained eleven input units, three hidden units, and two output units. Otherwise, the model uses the same activation and loss functions described previously. As with previous models, input consisted of a set of semantic features representing a verb, which the model categorized as either a double object dative response, or a prepositional dative response. Additionally, one input unit signaled when the transfer action being described was complete.

During the life experience phase, the model learned biases for 10 different dative verbs. The model received 200 epochs of life training, and saw each verb 10 times. The precise life experience biases learned for each verb can be seen in Table 20. Biases were taken either from Lin (2020), or estimated for verbs which do not have established norms.

Table 20: Verb Biases for Life Training

Verb	DO Bias	PD Bias
Throw	0%	100%
Offer	20%	80%
Bring	40%	60%
Slide	20%	80%
Give	50%	50%
Mail	20%	80%
Roll	20%	80%
Pass	10%	90%
Show	60%	40%
Toss	10%	90%

Table 21: Verb Representations for Thothathiri and Braiuca (2021)

Verb	Hand-Related	Caused Motion	Caused Possession	Sending	Deictic	Ballistic	Path-focused	Force	Recipient-focused	Specificity
Throw	1	1	1	0	0	1	1	1	0	0
Offer	1	0	0	0	0	0	0	0	1	0
Bring	1	1	1	1	1	0	0	0	1	0
Slide	0	1	0	1	0	0	1	0	0	0
Give	0	0	1	0	0	0	0	0	1	0
Mail	0	1	1	1	0	0	1	0	0	1
Roll	0	1	0	1	0	0	1	1	0	0
Pass	1	0	1	0	0	0	1	0	0	0
Show	0	0	0	0	0	0	0	0	1	0
Toss	1	1	1	0	0	1	1	0	0	0

During the experimental phase, the model saw the same lists used in Thothathiri and Braiuca (2021). The lists are reproduced below in Tables 22 and 23. Note that in Experiment 1, the majority of verbs occur in only one structure, while in Experiment 3 the majority of verbs occur equally in both structures. This creates a distribution with more alternation, as well as changing how well cues predict a particular syntactic category. In Experiment 1, both verbs and the completed-transfer cue are strong predictors of syntactic structure. In Experiment 3, the completed-transfer cue is a stronger predictor. Finally, after experiencing either the training used in Experiment 3 or the training used in Experiment 1, each verb was evaluated for its total bias toward the double object dative. A total of 50 individual models were trained to produce the results of Experiment 1, and 48 individual models were used to produce the results of Experiment 3. Details of the training phases can be seen below in Fig. 40. Finally, while the completed-transfer cue always perfectly predicted the use of the double object dative in the experimental training, multiple different cue structures were tried during the life experience phase. These life cues are explained in full in the results for each cue type.

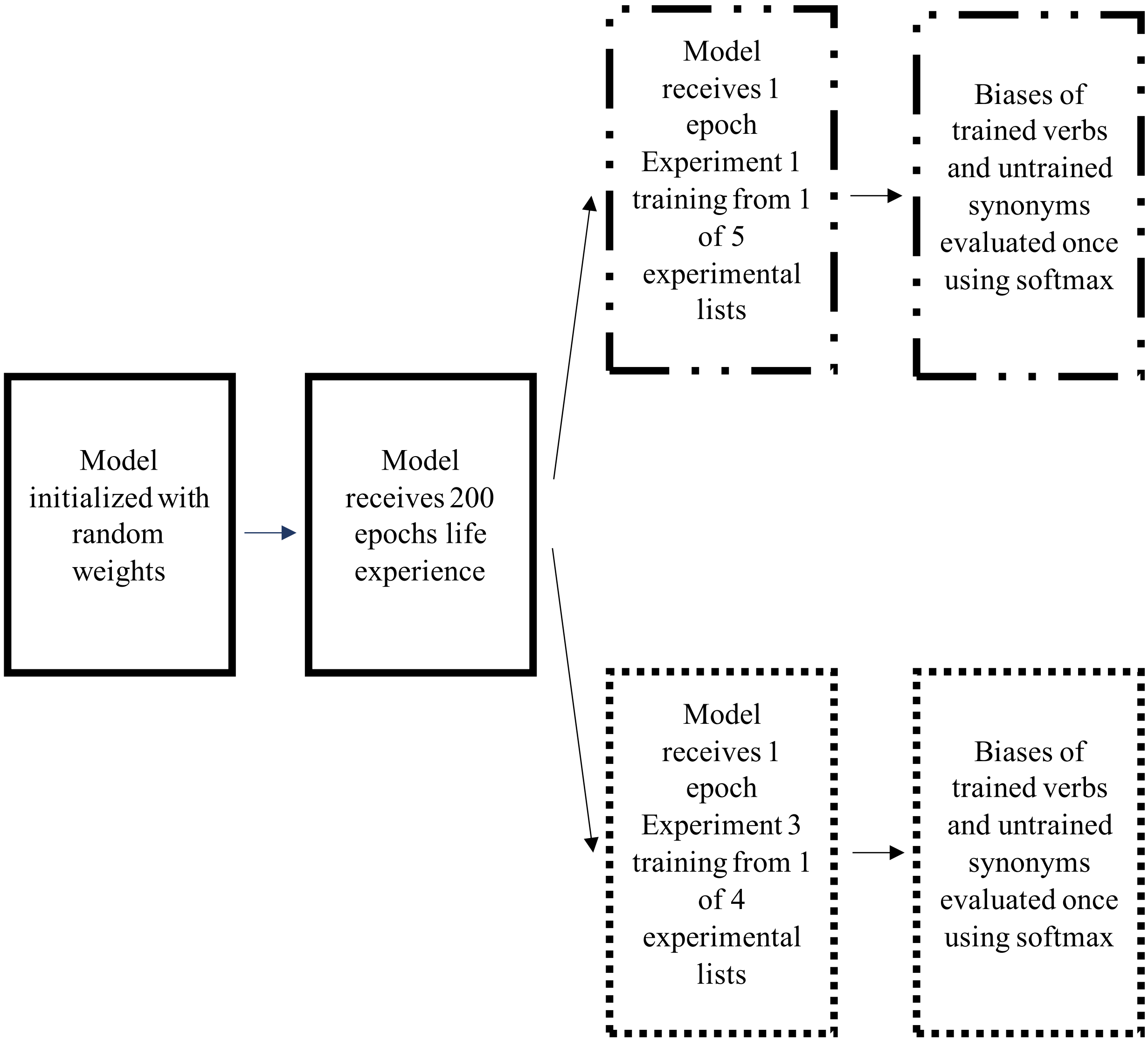

Figure 40: Life experience and experimental training for Thothathiri and Braiuca (2021) modeling

Table 22: Lists for Thothathiri and Braiuca (2021), Experiment 1

List	DO-trained verbs	Alternating verbs	PD-trained verbs
1	Throw, offer, bring, slide	Give, mail	Roll, pass, show, toss
2	Give, roll, show, toss	Throw, pass	Offer, bring, slide, mail
3	Mail, pass, show, toss	Offer, roll	Throw, bring, slide, give
4	Offer, slide, give, mail	Bring, toss	Throw, roll, pass, show
5	Throw, bring, roll, pass	Slide, show	Offer, give, mail, toss

Table 23: Lists for Thothathiri and Braiuca (2021), Experiment 3

List	DO-trained verb	Alternating verbs	PD-trained verb
1	Give	Throw, offer, bring, slide, mail, roll, pass, toss	Show
2	Bring	Throw, offer, slide, give, mail, roll, show, toss	Pass
3	Pass	Throw, offer, slide, give, mail, roll, show, toss	Bring
4	Show	Throw, offer, bring, slide, mail, roll, pass, toss	Give

7.2.2 Results

As a starting point, the first completed-transfer cue used was always equal to zero, which corresponds to a scenario where this cue is never used in English. Using this structure, the model seems to learn verb biases in both experimental conditions. In Experiment 1, DO-trained verbs are more likely to occur in the double-object dative than the alternating verbs (DO-trained DO-bias: 0.593, Alternating DO-bias: 0.502, $t(49)=-4.33$, $p<0.05$). Although the difference between the alternating verbs and the PD-trained verbs does not reach significance, the trend suggests that the PD-trained verbs are numerically less biased toward the double object dative than the alternating verbs (Alternating DO-bias: 0.502, PD-trained DO-bias: 0.465, t(49)=-1.74, p>0.05). Results for Experiment 1 can be seen in Fig. 41. This same pattern can be seen in Fig. 42, which shows the results for Experiment 3. In this case, the DO-trained verb is more biased toward the double object dative than the alternating verbs, and the alternating verbs are more biased toward the double object dative than the PD-trained verbs (DO-trained: 0.623, Alternating: 0.537, t(47)=-3.49, p<0.05; Alternating: 0.537, PD-trained: 0.412, t(47)=-4.17, p<0.05). Consequently, when the completed-transfer feature is always equal to zero in life experience, the model always learns verb biases. However, since the empirical results suggest that learning should be different between the two experiments, it appears that this cue contingency is not entirely correct.

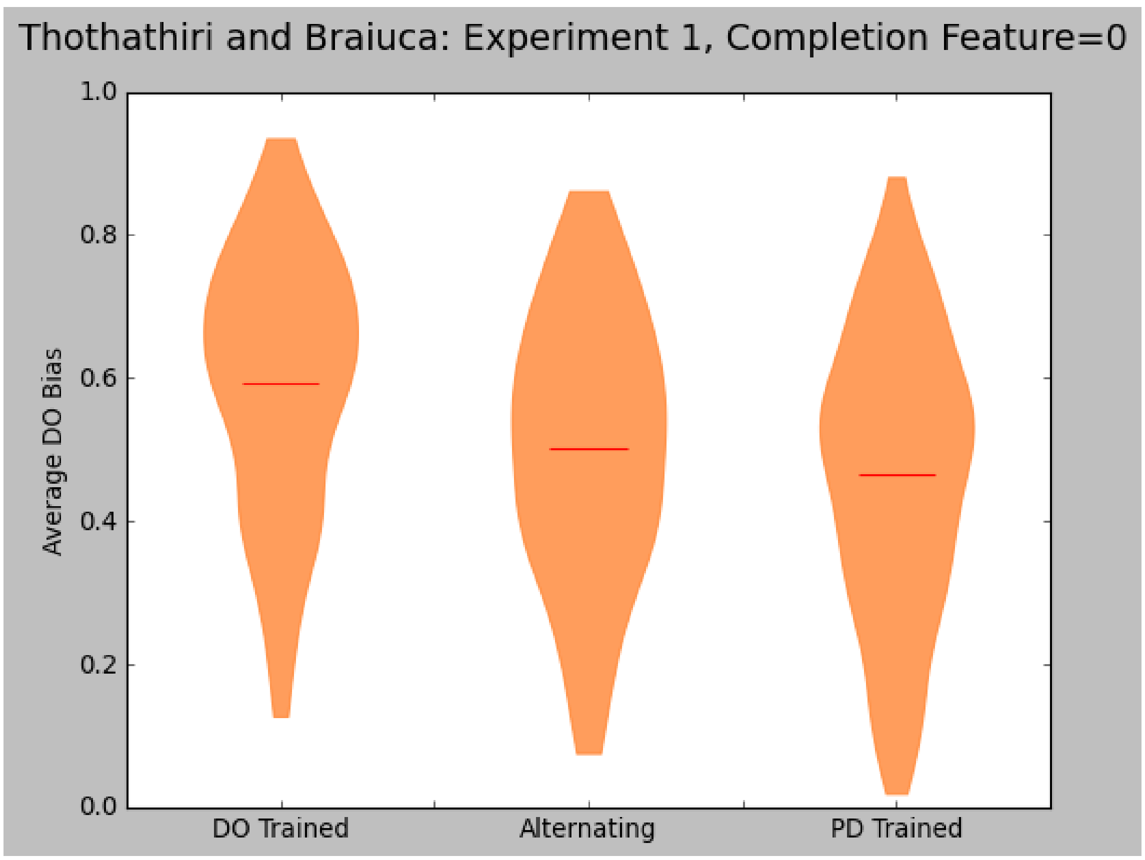

Figure 41: DO bias for each type of verb in Experiment 1 of Thothathiri and Braiuca when the completed-transfer cue in life experience is always equal to zero.

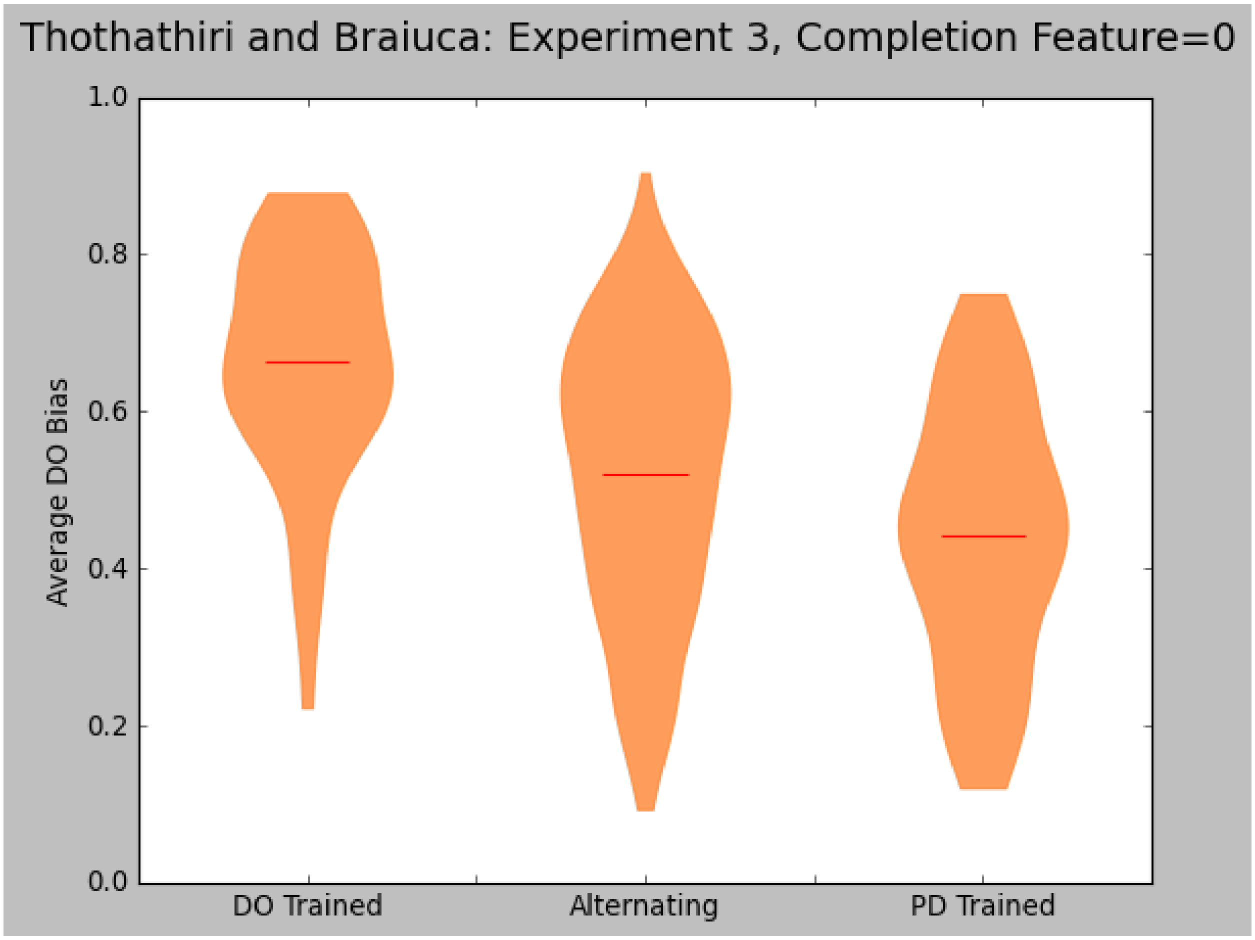

Figure 42: DO bias for each type of verb in Experiment 3 of Thothathiri and Braiuca when the completed-transfer cue in life experience is always equal to zero.

The next cue structure assumes that completed-transfer cue is perfectly predictive of the double object dative structure. In other words, the cue is always one when the verb should be categorized into the double object dative, and is always zero when the verb will appear in the prepositional dative. This cue structure totally eliminates verb bias learning in Experiment 1, which is visualized in Fig. 43 (DO-trained: 0.494, Alternating: 0.517, $t(49)=-0.30$, $p>0.05$); Alternating: 0.517, PD-trained: 0.504, $t(49)=0.45$, $p>0.05$). Additionally, while the DO-trained bias is higher than the alternating bias for Experiment 3, double object bias is no longer graded by training and consequently this study does not demonstrate verb bias learning (DO-trained: 0.64, Alternating: 0.49, $t(47)=-3.10$, $p<0.05$; Alternating: 0.49, PD-trained: 0.62, $t(47)=2.34$, p<0.05). Results are presented in Fig 44, and show visually that double object bias is approximately the same for both DO-trained and PD-trained verbs.

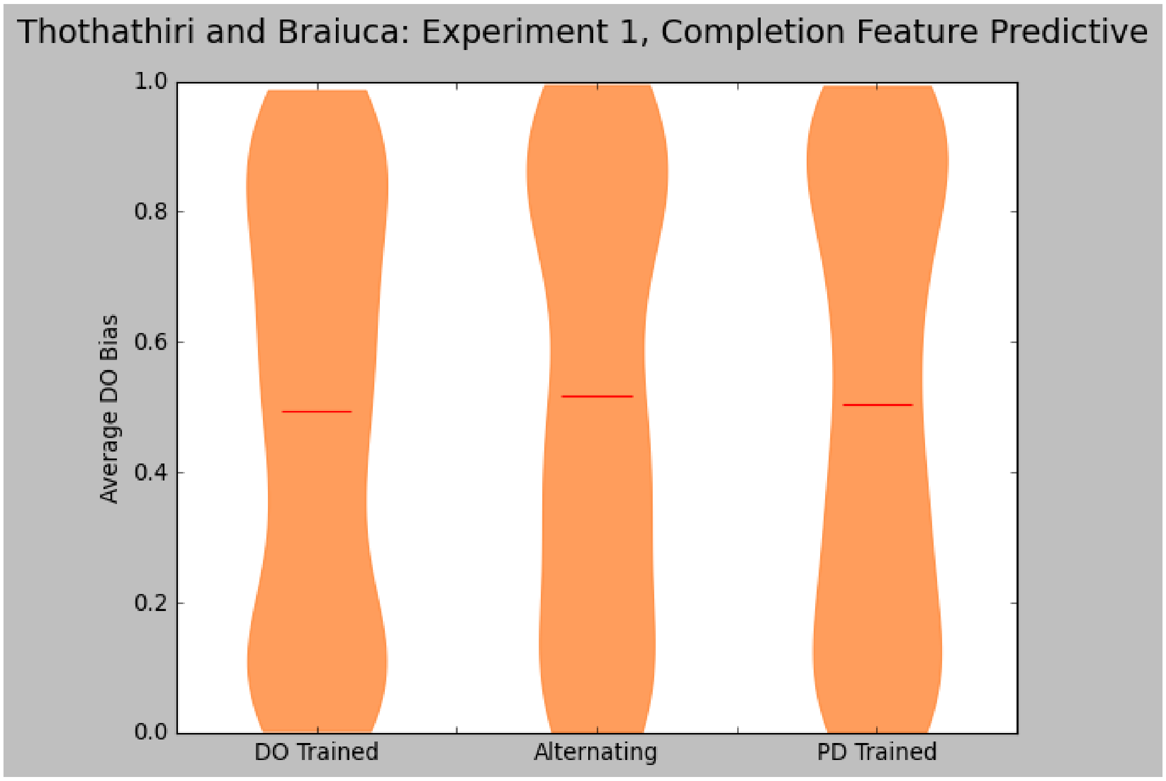

Figure 43: Results for Experiment 1 when completed-transfer cue is perfectly predictive of DO structure.

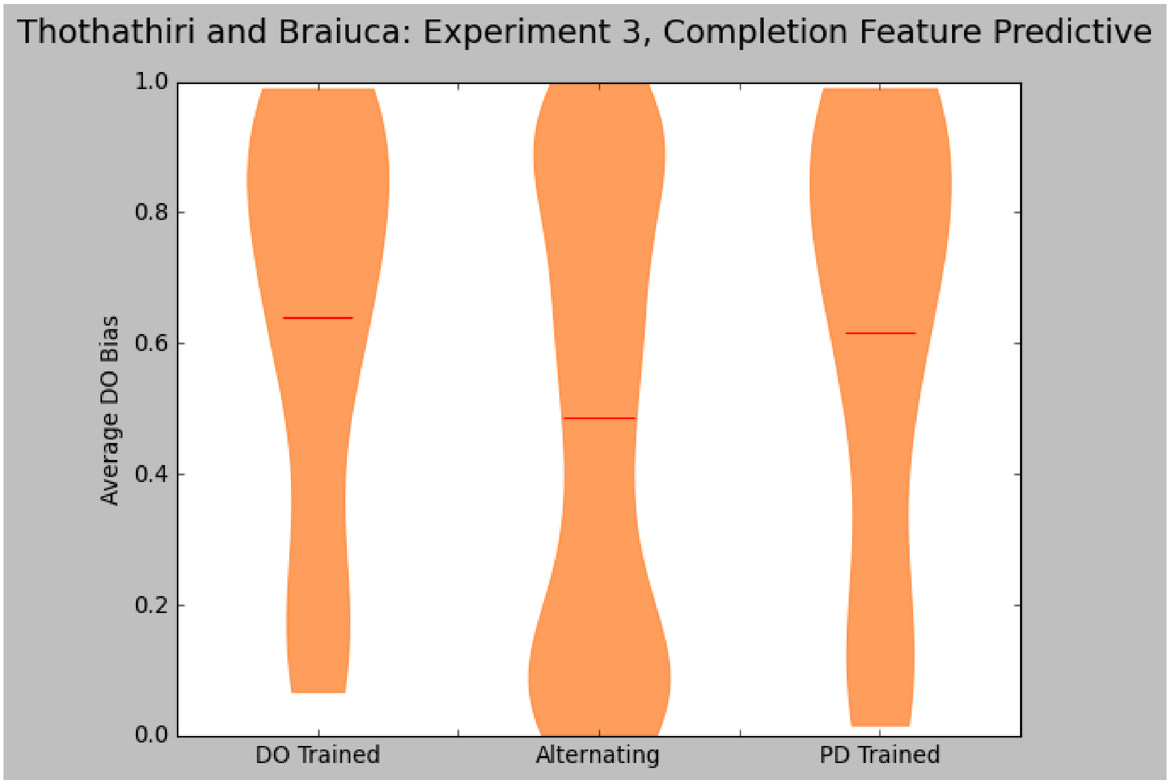

Figure 44: Results for Experiment 1 when completed-transfer cue is perfectly predictive of DO structure.

Based on these two analyses, it appears that both capture some aspects of the empirical results seen in Thothathiri and Braiuca (2021). When the completed-transfer feature is always 0, the results for Experiment 1 closely resemble the empirical results. Conversely, when the completed-transfer feature perfectly predicts the double object dative, it blocks verb bias learning and replicates the results of Experiment 3. Presumably, the completed-transfer feature that can successfully replicate the empirical results splits the difference between these two extremes.

The first compromise was to make the completed-transfer feature predict randomly. This feature took a value of either one or zero approximately half the time, but did not predict the use of either a double object dative or prepositional dative. This was intended to test a case where the feature was turned on sometimes during life experience, but did not block learning about verb representations by being predictive of the sentence structure. Each separate model received a unique random series of ones and zeroes to serve as its completed-transfer feature. However, this feature did not successfully replicate the empirical results from Thothathiri and Braiuca (2021). The differences between the conditions in Experiment 1 did not reach significance (DO-trained: 0.56, Alternating: 0.54, t(49)=-1.89, p>0.05; Alternating: 0.54, PD-trained: 0.49, t(49)=-0.55, p>0.05). Experiment 3 found a significant difference between the DO-trained and alternating conditions (DO-trained: 0.66, Alternating: 0.53, t(47)=-3.71, p<0.05), but not between the PD-trained and alternating conditions (Alternating: 0.53, PD-trained: 0.49, t(47)=-0.95, p>0.05). Additionally, although the results do not all reach significance, the numerical trend seen in Figs. 45 and 46 both most resemble the behavioral results found in Experiment 1.

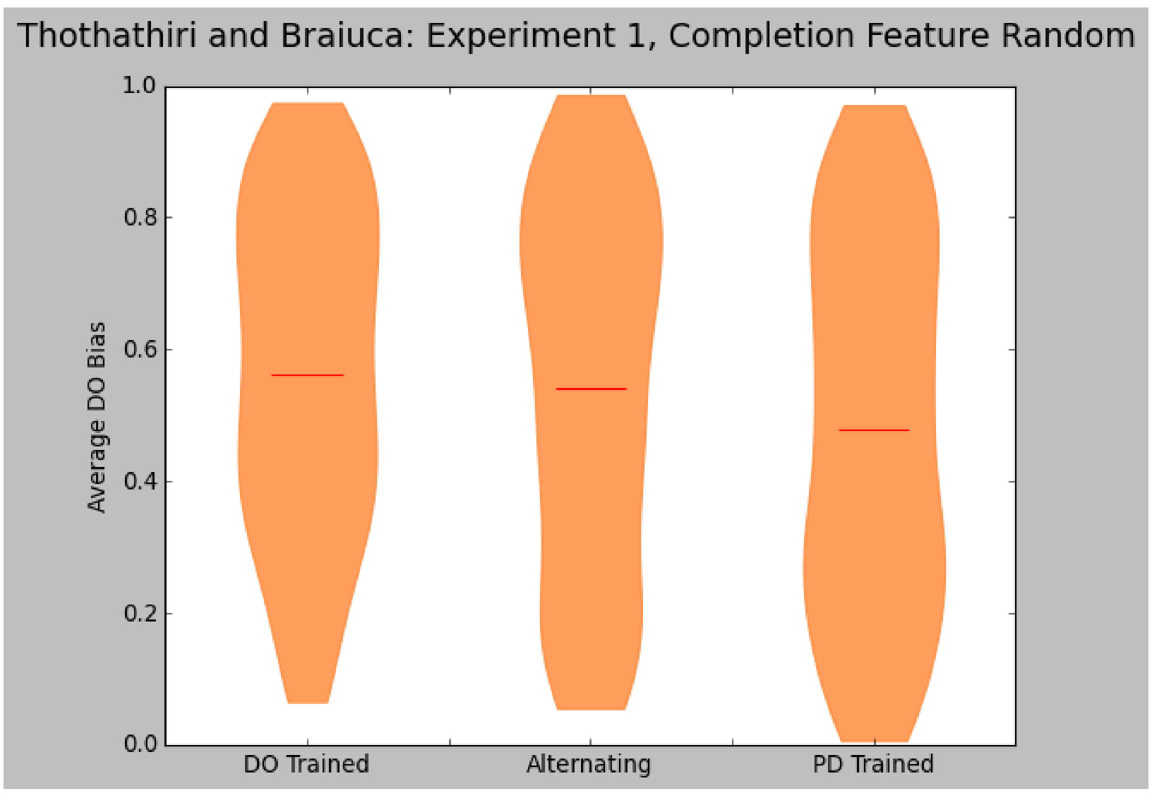

Figure 45: Model results for Experiment 1 when completion feature is randomly assigned to be either 1 or 0.

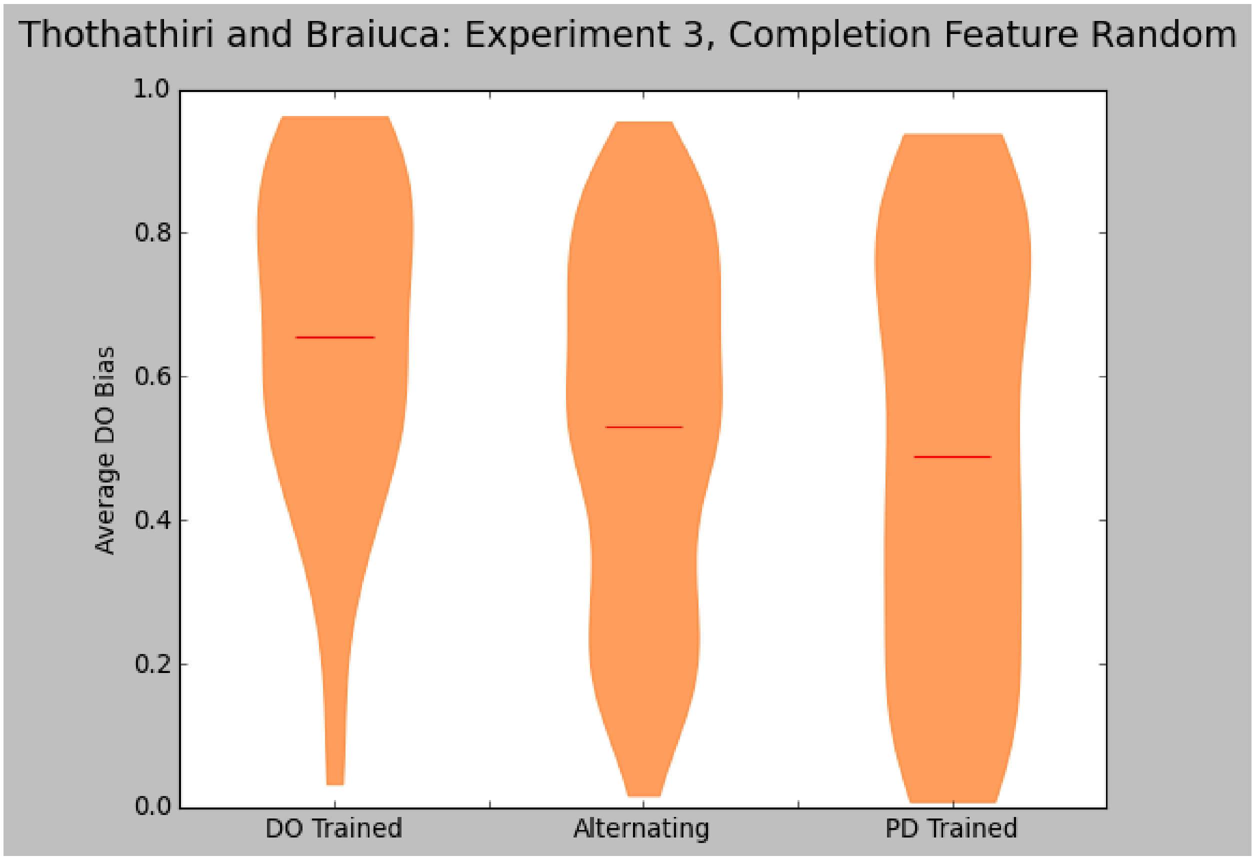

Figure 46: Model results for Experiment 3 when completion feature is randomly assigned to be either 1 or 0.

Based on these results, the random completion feature is not a good approximation of completion information in real language. Instead, the completion feature must do a better job of splitting the difference between being always zero and being perfectly predictive. In other words, this feature should be close to zero, and somewhat predictive of sentence structure. To fulfill these requirements, a new kind of random feature was created. This feature was zero most of the time, but was slightly more likely to be one when the structure was a DO. The process for creating this feature is demonstrated in Fig. 47. The process began with creating two series of random numbers that fell between zero and one. Since the mean of these numbers is 0.5, subtracting a number between 0 and 0.5 will shift the mean of this series downward. After rounding, this will result in a feature that is primarily zero, with an occasional one. The feature that produced completed-action values for the double object dative was the result of subtracting 0.43 before rounding, and the completed-action values for the prepositional dative were created by subtracting 0.49. Consequently, the DO feature is slightly more likely to be one than the PD feature. Finally, the completion feature is created by drawing elements from these two series that correspond to the correct sentence structure. For example, if a verb received seven PD trials and then three DO trials, the first seven completion features would be drawn from the PD series, and the final three would be drawn from the DO series. Overall, this creates a feature where the completed-action feature is activated infrequently, but when it is, it is more likely to predict the double object dative than the prepositional dative. Each model received a different random series of completion features for life training. The mean value of the completion feature series was 0.02.

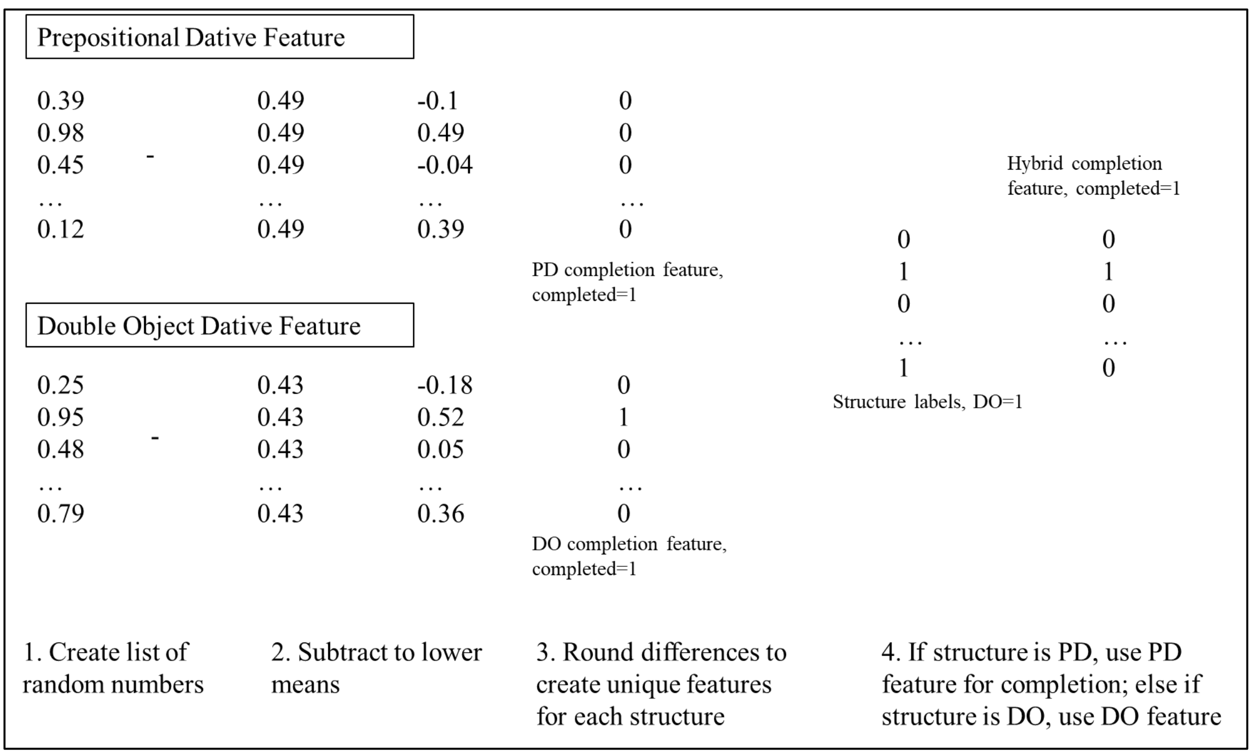

Figure 47: Steps used to create a completed-action feature that is slightly predictive of sentence structure.

The model results of the studies that used this feature can be seen in Figs. 48 and 49 below. For Experiment 1, the DO-trained verbs were significantly different from the PD-trained verbs (DO-trained: 0.59, PD-trained: 0.47, $t(49)=-5.36$, $p<0.05$). The PD-trained verbs were significantly different from the alternating verbs (Alternating: 0.55, PD-trained: 0.47, $t(49)=-2.63$, $p<0.05$). However, no significant difference was found between the DO-trained and alternating verbs (DO-trained: 0.59, Alternating: 0.55, $t(49)=-1.07$, $p<0.05$). Finally, the trend in the data visually resembles the linear trend reported in Thothathiri and Braiuca (2021).

For Experiment 3, the DO-trained verb was not more likely to be categorized as a double object than PD-trained verb (DO-trained: 0.53, PD-trained: 0.51, $t(47)=-0.92$, $p>0.05$). Additionally, no difference was found between the alternating and PD-trained verbs (Alternating: 0.49, PD-trained: 0.51, $t(47)=0.49$, $p>0.05$), or between the DO-trained and alternating verbs (DO-trained: 0.53, Alternating: 0.49, $t(47)=-1.44$, $p>0.05$). Additionally, this trend is visually different from Experiment 1, and appears to be similar to the change from a linear trend to a lack of an effect seen in Thothathiri and Braiuca (2021).

In order to directly test the difference between these two model experiments, the difference between the DO and PD-trained verbs was compared using a Welch's independent samples t-test. The differences in Experiment 1 were consistently larger than those in Experiment 3 (Experiment 1: 0.11, Experiment 3: 0.02, $t(90.76) = 2.90$, $p<0.05$). This suggests that the trends in Experiments 1 and 3 represent different amounts of verb bias learning between the two experiments.

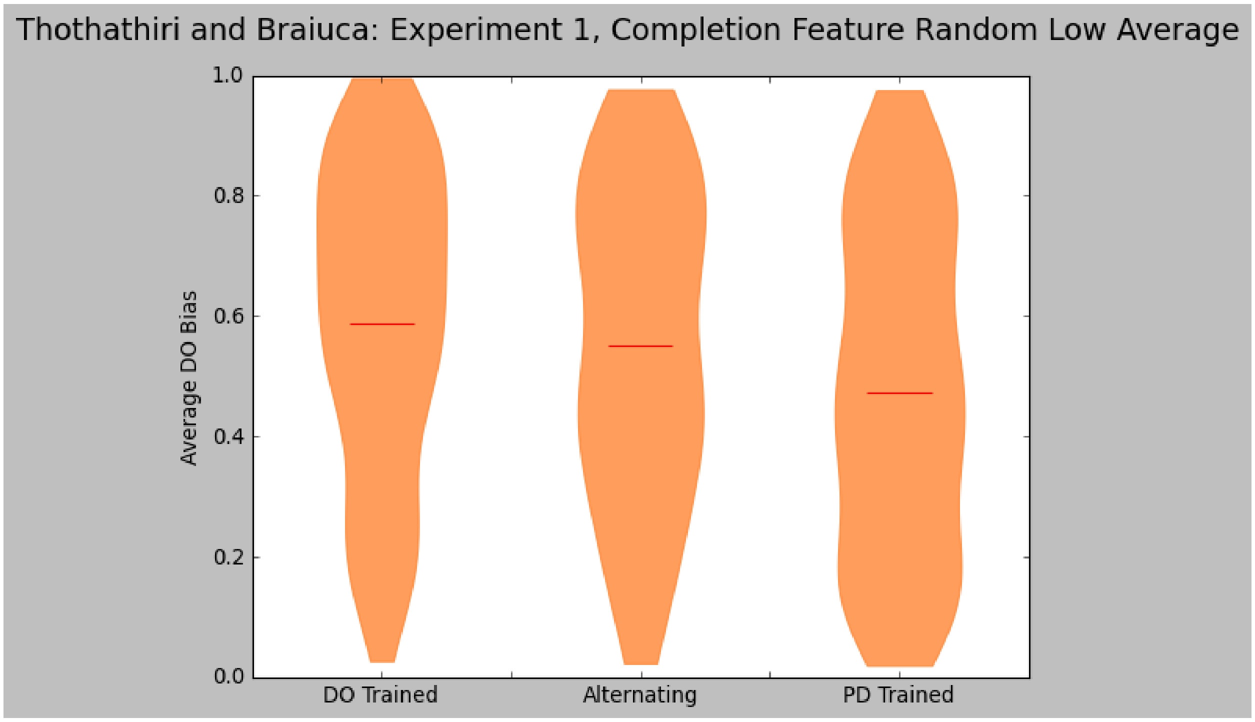

Figure 48: Model results for Experiment 1 of Thothathiri and Braiuca (2021) using a completed-action feature with a low average that varied depending on whether a PD or DO structure was used in training.

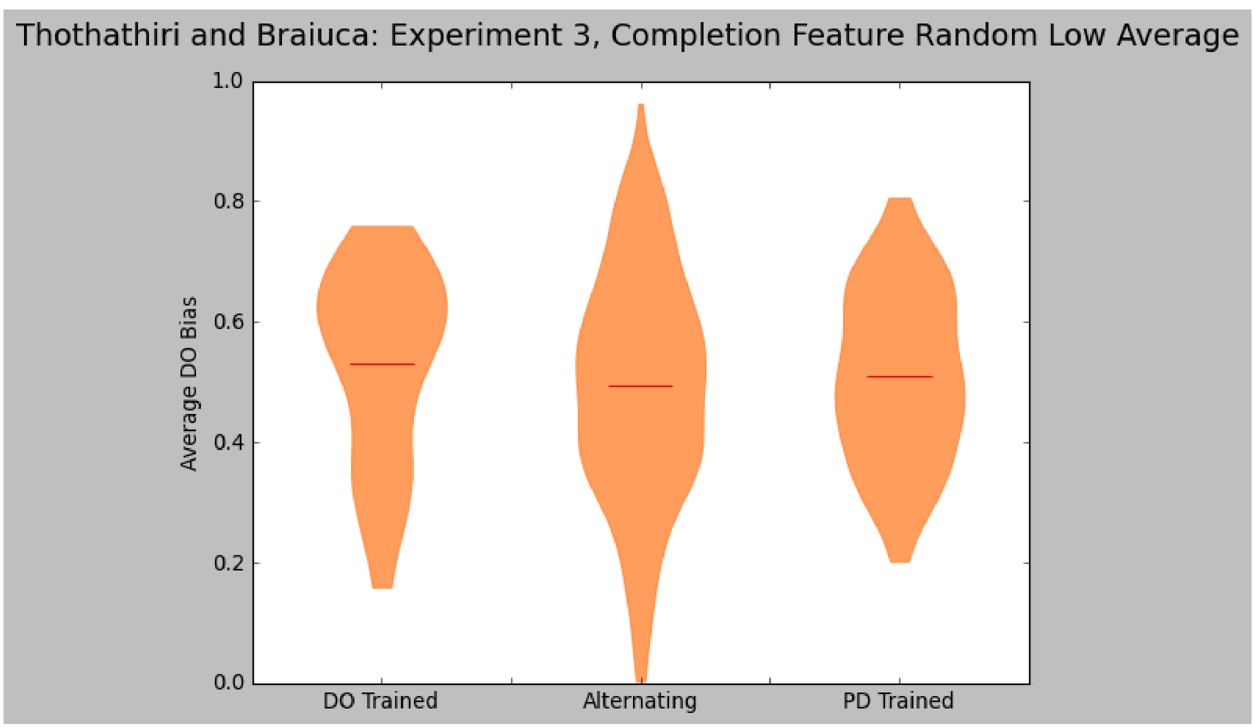

Figure 49: Model results for Experiment 3 of Thothathiri and Braiuca (2021) using a completed-action feature with a low average that varied depending on whether a PD or DO structure was used in training.

7.2.3 Conclusions

Replicating the results of Thothathiri and Braiuca (2021) depended on tuning the model's life experience correctly. Specifically, it required finding a completed-action feature that was both slightly predictive of the structure of the sentence, but also was not activated frequently. Using this feature, the verb bias training effect is reduced in Experiment 3 compared to Experiment 1. These changes resemble the empirical results found in Thothathiri and Braiuca, and suggest that the model has captured something meaningful about the differences in the two experiments.

One objection to these findings is that finding the same results as Thothathiri and Braiuca (2021) required significant hand-turning of the model. An ideal model would be able to switch between learning that the completion feature is irrelevant in Experiment 1, and learning that it is perfectly predictive in Experiment 3. The model that can replicate both results does so seemingly by finding a sort of halfway point in the distribution of the life experience feature that allows the model to approximate both results. Improving the model architecture could automate this switch, and avoid the need to manually search for the halfway point. Specific changes will be further discussed in the general conclusions. However, another important aspect of successfully replicating this experiment is gathering more empirical data. In comparisons between their experiments, Thothathiri and Braiuca find that they cannot definitively rule out an effect of verb bias in Experiment 3, and suggest that verb-specific cues may never be fully discounted by speakers. Consequently, continued experimentation is needed to more fully understand how these cues compete, and subsequently to guide future modeling.

7.3 GENERAL CONCLUSIONS

Overall, both Lin (2020) and Thothathiri and Braiuca (2021) suggest that verb bias learning is more complex than learning weights between a verb and a structure. Rather, it is both learning that individual verbs are good predictors of structure, and then learning what structures to associate with those verbs. Consequently, the models presented in this chapter and previous chapters should do a good job accounting for behavioral results when verbs are generally a good predictor. Presumably, this is the case for the dative findings modeled in Chapters 2 and 5. It is also the case that when verbs are generally bad predictors and that learning can be blocked by a single, more predictive cue, then the model can still account for a pattern of behavioral results. This is the outcome seen in Chapter 6, where a patient-first cue blocked verb bias learning of transitive verbs. However, when it becomes important for the model to change how predictive a cue is based on experimental experience, the model no longer performs as well. Although careful tuning allowed it to replicate the Thothathiri and Braiuca results to some degree, it unequivocally failed to replicate the results of Lin (2020). Together, these failures make a case for specific improvements to model architecture, with broader implications for where verb biases exist in a larger framework of language statistics.

As briefly discussed in the conclusions for Study 1, one possible solution to this problem is to implement an attention layer in the model. In essence, an attention mechanism allows a model to weight some parts of its inputs more heavily than others (e.g., Chorowski, Bahdanau, Cho, and Bengio, 2014; Karmakar, Teng, and Lu, 2021). This concept has many applications – for instance, learning to identify groups of words in sentences that are important for successful machine translation (Chorowski, Bahdanau, Cho, and Bengio, 2014). There are many mechanisms that implement attention (Karmakar, Teng, and Lu, 2021), and many kinds of models that can use it, including feed-forward models similar to the type used in this chapter (e.g., Raffel and Ellis, 2015). For learning associations between cues and structures, an attention mechanism would essentially implement the kind of switching hypothesized by both Lin (2020) and Thothathiri and Braiuca (2021). In the first case, weight would shift from the verb to the verb class, and in the second, from the verb to the semantics of the sentence.

In a sense, this process brings the neural net model closer to the multi-level learning that occurs in some Bayesian models of verb learning (e.g., Perfors, Tenenbaum, and Wonnacott, 2010; Barak, Fazly, and Stevenson, 2014). However, rather than creating a model with a priori levels, a model with an attention mechanism could learn cues from the input on its own. The precise details of this depend not only on how the attention mechanism is implemented, but also on the type of input the model receives. However, it is not difficult to imagine a model that receives a distributed input vector that contains information like verb semantics, event semantics, verb class, and production constraints like givenness or focus. With the appropriate attention mechanism, the model should be able to focus on predictive aspects of this input, and ignore any that are not predictive. Additionally, the model should be able to redirect attention with appropriate experimental training, potentially including the type of training used in both Lin (2020) and Thothathiri and Braiuca (2021).

Model performance on the results of Lin (2020) and Thothathiri and Braiuca (2021) suggests the current model is not sufficient to explain these types of cue-switching findings. However, the addition of a mechanism like attention may allow the model to better account for these results, and has the potential to generate new predictions about learning. So far, cues like completed transfer or verb class have been identified based on a priori theories about what kinds of cues participants are likely to use in the process of learning. With enough training and the right input, an unsupervised attention mechanism could identify novel cues that contribute to distributional learning.

Additionally, there is evidence from other arenas of language production that suggests that some distributions are easier to learn than others. One example of this is that participants are able to quickly learn new phonotactic rules, which govern where sounds can occur in a syllable, so long as these rules only place restrictions based on position within the syllable. For instance, if the rule says that /f/ can only occur at the beginning of a syllable, this rule is acquired rapidly (Dell, Reed, Adams, and Meyer, 2000). However, if the rule is instead that /f/ occurs at the beginning of syllables only when the vowel is /e/, then this rule requires a period of sleep consolidation to learn (Gaskell, Warker, Lindsay, Frost, Guest, Snowdon, and Stackhouse, 2014; Anderson and Dell, 2018). Finally, some rules cannot be learned – for instance, English speakers show no evidence of learning when a phonotactic rule is conditioned on a specific type of tone (Bian and Dell, 2020). Learning these distributions is presumably based on life experience with phonotactic rules that English speakers bring to these experiments. Given broad similarities between learning in language production at multiple levels, it seems likely that similar learnability constraints apply to learning verb distributions. A combination of experimental research and cognitive modeling could begin to explicitly model how life experience allows speakers to freely switch between some cues, while being unable to condition verb learning distributions on others.

REFERENCES

Ambridge, B., & Blything, R. P. (2016). A connectionist model of the retreat from verb argument structure overgeneralization. *Journal of child language*, *43*(6), 1245-1276. doi:10.1017/S0305000915000586.

Ambridge, B., Pine, J.M., & Rowland, C.F. (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics*, *22*(2), 303-323. https://doi.org/10.1515/cogl.2011.012.

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb semantics, entrenchment, and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language*, *88*(1), 45-81. http://www.jstor.org/stable/41348883.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*(1), 87-129. https://doi.org/10.1016/j.cognition.2006.12.015.

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences*, *115*(14), 3617-3622. https://doi.org/10.1073/pnas.1721107115.

Anible, B., Twitchell, P., Waters, G. S., Dussias, P. E., Piñar, P., & Morford, J. P. (2015). Sensitivity to verb bias in American Sign Language–English bilinguals. *Journal of deaf studies and deaf education*, *20*(3), 215-228. https://doi.org/10.1093/deafed/env007.

Chorowski, J., Bahdanau, D., Cho, K., & Bengio, Y. (2014). End-to-end continuous speech recognition using attention-based recurrent nn: First results. In *NIPS 2014 Workshop on Deep Learning, December 2014.* https://doi.org/10.48550/arXiv.1412.1602.

Barak, L., Fazly, A., & Stevenson, S. (2014, June). Learning verb classes in an incremental model. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics*, 37-45. http://www.aclweb.org/anthology/W/W14/W14-2005.

Barak, L., Goldberg, A., & Stevenson, S. (2016, November). Comparing computational cognitive models of generalization in a language acquisition task. In *Proceedings of the 2016 conference on Empirical Methods in Natural Language Processing*, 96-106. https://doi.org/10.18653/v1/D16-1010.

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming?. *Cognition*, *114*(3), 455-461. https://doi.org/10.1016/j.cognition.2009.11.005.

Bian, Y., & Dell, G. S. (2020). Novel stress phonotactics are learnable by English speakers: Novel tone phonotactics are not. *Memory & Cognition*, *48*(2), 176-187. https://doi.org/10.3758/s13421-019-01000-9.

Bock, J.K. (1986). Syntactic Persistence in Language Production. *Cognitive Psychology*, *18*(3), 355-387. https://doi.org/10.1016/0010-0285(86)90004-6.

Braine, D.S., & Brooks, P.J. (1995). Verb Argument Structure and the Problem of Avoiding an Overgeneral Grammar. In M. Tomasello and W.E. Merriman (Eds.), *Beyond Names for Things*. Pp. 353-376. Psychology Press. https://doi-org.proxy2.library.illinois.edu/10.4324/9781315806860.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69-94). KNAW. http://hdl.handle.net/11858/00-001M-0000-0013-1A32-0.

Brooks, P.J., and Tomasello, M. (1999). How children constrain their argument structure productions. *Language*, *75*(4), 720-738. https://doi.org/10.2307/417731.

Cantwell, G., Crossley, M.J., & Ashby, F.G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin and Review*, *22*, 1598-1613. https://doi.org/10.3758/s13423-015-0827-2.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609-651. https://doi.org/10.1207/s15516709cog2605_3.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review*, *113*(2), 234-272. https://doi.org/10.1037/0033-295X.113.2.234.

Coyle, J. M., & Kaschak, M. P. (2008). Patterns of experience with verbs affect long-term cumulative structural priming. *Psychonomic bulletin & review*, *15*(5), 967-970. https://doi.org/10.3758/PBR.15.5.967.

Cureton, R. D. (1979). The exceptions to passive in English: a pragmatic hypothesis. *Studies in the Linguistic Sciences*, *9*(2), 39-53. https://www.researchgate.net/profile/Richard-Cureton/publication/331998083_The_Exceptions_to_Passive_in_English/links/5d30d1e892851cf440900c5a/The-Exceptions-to-Passive-in-English.pdf.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47(1), 1-12. https://doi.org/10.3758/s13428-014-0458-y.

DeDe, G. (2013). Effects of verb bias and syntactic ambiguity on reading in people with aphasia. *Aphasiology*, *27*(12), 1408-1425. https://doi.org/10.1080/02687038.2013.843151.

Dell, G.S., Kelley, A.C., Hwang, S., & Bian, Y. (2021). The adaptable speaker: A theory of implicit learning in language production. *Psychological Review, 128*(3), 446–487. https://doi.org/10.1037/rev0000275.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1355–1367. https://doi.org/10.1037/0278-7393.26.6.1355.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (Second). John Wiley & Sons, Inc.

Dussias, P. E., & Scaltz, T. R. C. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta psychologica*, *128*(3), 501-513. https://doi.org/10.1016/j.actpsy.2007.09.004.

Elman, J.L. (1990). Finding Structure in Time. *Cognitive Science, 14*, 179-211. https://doi.org/10.1016/0364-0213(90)90002-E

Fellows, L. K., & Farah, M.J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain*, 126, pp. 1830-1837. https://doi.org/10.1093/brain/awg180.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*(6), 715-736. https://doi.org/10.1006/jmla.1994.1034.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555. https://doi.org/10.1037/0278-7393.16.4.555.

Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors?. *Quarterly Journal of Experimental Psychology*, *66*(8), 1548-1571. https://doi.org/10.1080/17470218.2012.753924.

Fisher, C., Jin, K. S., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in cognitive science*, *12*(1), 48-77. https://doi.org/10.1111/tops.12447.

Frank, M., & Claus, E. (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision-making, and reversal. *Psychological Review, 113*(2), pp. 300-326. https://doi.org/10.1037/0033-295X.113.2.300.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210. https://doi.org/10.1016/0010-0285(82)90008-1.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 748-775. https://www.jstor.org/stable/4489780.

Gahl, S., Garnsey, S. M., Fisher, C., & Matzen, L. (2006). That sounds unlikely": Syntactic probabilities affect pronunciation. In *Proceedings of the 27th meeting of the Cognitive Science Society*. https://escholarship.org/uc/item/1h9879s9.

Gahl, S., & Menn, L. (2016). Usage-based approaches to aphasia. *Aphasiology*, *30*(11), 1361-1377. https://doi.org/10.1080/02687038.2016.1140120.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of memory and language*, *37*(1), 58-93. https://doi.org/10.1006/jmla.1997.2512.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep underpins the plasticity of language production. *Psychological Science*, *25*(7), 1457-1465. https://doi.org/10.1177/0956797614535937.

Goldberg, A.E. (1995). *Constructions: A construction grammar approach to argument structure.* University of Chicago Press.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178-186. https://doi.org/10.1016/S1364-6613(00)01467-4.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, *48*(2), 281-303. https://doi.org/10.1016/S0749-596X(02)00516-8.

Hare, M., McRae, K., & Elman, J. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, *19*(2), 181-224. https://doi.org/10.1080/01690960344000152.

Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, *33*(4), 610-628. https://doi.org/10.1111/j.1551-6709.2009.01027.x.

Hawkins, R. D., Yamakoshi, T., Griffiths, T. L., & Goldberg, A. E. (2020). Investigating representations of verb bias in neural language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 4653-4663. https://doi.org/10.18653/v1/2020.emnlp-main.376.

Hoffmann, F. K., Perkins, M. J., & Calvin, A. D. (1956). The effect of problem difficulty on discrimination reversal. *Journal of comparative and physiological psychology*, *49*(6), 547-548. https://doi.org/10.1037/h0041304.

Izquierdo, A., Brigman, J.L., Radke, A.K., Rudebeck, P.H., & Holmes, A. (2017). The Neural Basis of Reversal Learning: An Updated Perspective. *Neuroscience,* 345, 12-26. https://doi.org/10.1016/j.neuroscience.2016.03.021.

Jacobs, C.L., Cho, S., Watson, D.G. (2019). Self-Priming in Production: Evidence for a Hybrid Model of Syntactic Priming. *Cognitive Science*, *43*(7), e12749. https://doi.org/10.1111/cogs/12749.

Karmakar, P., Teng, S. W., & Lu, G. (2021). *Thank you for attention: a survey on attention-based artificial neural networks for automatic speech recognition.* arXiv. https://doi.org/10.48550/arXiv.2102.07259

Kelley, A.C. (2019). *Investigating the mechanisms of verb bias learning*. [Master's thesis]. University of Illinois at Urbana-Champaign. IDEALS. http://hdl.handle.net/2142/105091.

Kendler, T.S., & Kendler, H.H. (1970). An Ontogeny of Optional Shift Behavior. *Child Development*, 41(1), pp. 1-27. https://www.jstor.org/stable/1127386.

Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*(2), 93-107. https://doi.org/10.1016/j.cogdev.2006.01.006.

Kidd, E., Lieven, E. V., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*(1), 132-142. https://doi.org/10.1016/j.langsci.2009.05.002.

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(1), 222-234. https://doi.org/10.1017/S0305000909990316.

Kruschke, J.K. (1990). ALCOVE: A connectionist model of human category learning. *Advances in Neural Information Processing Systems*, *3.* 649-655. https://proceedings.neurips.cc/paper/1990/hash/8fe0093bb30d6f8c31474bd0764e6ac0-Abstract.html

Kruschke, J.K. (1996). Dimensional Relevance Shifts in Category Learning. *Connection Science*, *8*(2), pp. 225-248. https://doi.org/10.1080/095400996116893.

Lee, E. K., Lu, D. H. Y., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761-775. https://doi.org/10.1017/S1366728912000776.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation.* The University of Chicago Press.

Levin, B. (2015). Semantics and pragmatics of argument alternations. *Annu. Rev. Linguist.*, *1*(1), 63-83. https://doi.org/10.1146/annurev-linguist-030514-125141.

Lin, Y. (2020). *Linking verbs to syntax.* [Doctoral dissertation, University of Illinois at Urbana-Champaign.] IDEALS. http://hdl.handle.net/2142/109327

Lin, Y., & Fisher, C. (2017). Error-based learning: A mechanism for linking verbs to syntax. In G. Gunzelman, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the 39th*

Annual Conference of the Cognitive Science Society, 754-759. Austin, TX: Cognitive Science Society. https://cogsci.mindmodeling.org/2017/papers/0151/paper0151.pdf.

McDaniel, J.W. (1969). Reversal learning in the sow bug. *Psychonomic Science*, *16*(5), pp. 261-262. https://doi.org/10.3758/BF03332678.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt and J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, 56-61. https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf.

Mitchell, D. C. (1989). Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes*, *4*(3-4), SI123-SI154. https://doi.org/10.1080/01690968908406366.

Murray, E.A., & Gaffan, D. (2006). Prospective Memory in the Formation of Learning Sets by Rhesus Monkeys (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), pp. 87-90. https://doi.org/10.1037/0097-7403.32.1.87.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, *18*(7), 1098-1111. https://doi.org/10.1162/jocn.2006.18.7.1098.

Olson, D.R., & Filby, N. (1972). On the Comprehension of Active and Passive Sentences. *Cognitive Psychology*, *3*(3), 361-381. https://doi.org/10.1016/0010-0285(72)90013-8.

The Pandas Development Team. (2020). pandas-dev/pandas: Pandas [Computer software]. Zenodo. Retrieved from: https://doi.org/10.5281/zenodo.3509134.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32(32), 2674-2679. https://escholarship.org/content/qt186313ch/qt186313ch.

Perek, F., & Goldberg, A. E. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language*, *84*, 108-127. https://doi.org/10.1016/j.jml.2015.04.006.

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276-293. https://doi.org/10.1016/j.cognition.2017.06.019.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of child language*, *37*(3), 607-642. https://doi.org/10.1017/S0305000910000012.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1-15. https://doi.org/10.1016/j.jml.2014.12.002.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of memory and Language*, *38*(3), 265-282. https://doi.org/10.1006/jmla.1997.2546.

Qi, Z., Yuan, S., & Fisher, C. (2011). Where does Verb Bias Come From? Experience with Particular Verbs Affects Online Sentence Processing. In N. Danis, K. Mesh, and H. Sung (Eds), *BUCLD 35: Proceedings of the 35th Annual Boston University Conference on Language Development*. http://web.mit.edu/zqi/www/uploads/1/4/8/9/14891652/bucld_final.pdf.

Qian, Z., Lee, E. K., Lu, D. H. Y., & Garnsey, S. M. (2019). Native and non-native (L1-Mandarin) speakers of English differ in online use of verb-based cues about sentence structure. *Bilingualism: Language and Cognition*, *22*(5), 897-911. https://doi.org/10.1017/S1366728918001001.

R Core Team. (2019). R: A language and environment for statistical computing [Computer Program]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.R-project.org/.

Raffel, C., & Ellis, D. P. (2015). *Feed-forward networks with attention can solve some long-term memory problems.* arXiv. https://doi.org/10.48550/arXiv.1512.08756.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914. https://doi.org/10.1002/wcs.78.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781. https://doi.org/10.1037/xlm0000341.

Ryskin, R., Qi, Z., Covington, N. V., Duff, M., & Brown-Schmidt, S. (2018). Knowledge and learning of verb biases in amnesia. *Brain and language*, *180*, 62-83. https://doi.org/10.1016/j.bandl.2018.04.003.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), p. 1926-1928. https://doi.org/10.1126/science.274.5294.1926.

Sanders, B. (1971). Factors affecting reversal and nonreversal shifts in rats and children. *Journal of Comparative and Physiological Psychology, 74*(2), pp. 192-202. https://doi.org/10.1037/h0030336.

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and cognitive processes*, *24*(6), 777-803. https://doi.org/10.1080/01690960802573236.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, *49*(3), 238-299. https://doi.org/10.1016/j.cogpsych.2004.03.001.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of child language*, *28*(1), 127-152. https://doi.org/10.1017/S0305000900004608.

Thompson, D., Ling, S.P., Myachykov, A., Ferreira, F., & Scheepers, C. (2013). Patient-related constraints on get- and be-passive uses in English: Evidence from paraphrasing. *Frontiers in Psychology*, *4*, 848. https://doi.org/10.3389/fpsyg.2013.00848.

Thothathiri, M., & Braiuca, M. C. (2021). Distributional learning in English: The effect of verb-specific biases and verb-general semantic mappings on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 113. https://doi.org/10.1037/xlm0000814.

Thothathiri, M., Evans, D. G., & Poudel, S. (2017). Verb bias and verb-specific competition effects on sentence production. *PloS one*, *12*(7), e0180580. https://doi.org/10.1371/journal.pone.0180580.

Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, *7*, 404. https://doi.org/10.3389/fpsyg.2016.00404.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147-165. https://doi.org/10.1515/LANGCOG.2009.008.

Torch Contributors. (2019). *CrossEntropyLoss*. Pytorch.org. https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html#torch.nn.CrossEntropyLoss.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134. https://doi.org/10.1016/S0010-0277(99)00032-3.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental psychology: Learning, memory, and Cognition*, *19*(3), 528-553. https://doi.org/10.1037/0278-7393.19.3.528.

Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, *73*, 41-71. https://doi.org/10.1016/j.cogpsych.2014.05.001.

Twomey, K. E., Chang, F., & Ambridge, B. (2016). Lexical distributional cues, but not situational cues, are readily used to learn abstract locative verb-structure associations. *Cognition*, *153,* 124-139. https://doi.org/10.1016/j.cognition.2016.05.001.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive psychology*, *56*(3), 165-209. https://doi.org/10.1016/j.cogpsych.2007.04.002.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1-14. https://doi.org/10.1016/j.jml.2011.03.001.

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?" Two-year-olds learn combinatorial facts about verbs by listening. *Psychological science*, *20*(5), 619-626. https://doi.org/10.1111/j.1467-9280.2009.02341.x.