YOUNG CHILDREN'S USE OF WORKING MEMORY FOR PRODUCING UNFAMILIAR SENTENCE STRUCTURES

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

ERYN ADAMS

Dr. Nelson Cowan, Dissertation Supervisor

JULY 2021

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

YOUNG CHILDREN'S USE OF WORKING MEMORY FOR PRODUCING UNFAMILIAR SENTENCE STRUCTURES

presented by Eryn Adams,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Nelson Cowan

Professor Jeffrey Johnson

Professor Yuyan Luo

Professor Elizabeth Kelley

DEDICATION

I dedicate my work over the last five years first to my family. I cannot express how the support of my parents, Jeffery and Sheri Adams, carried me through the mountains and valleys of this journey. Thank you for answering every phone call and pushing me through. Thank you to my brother, Jeff Jr., for the well-timed funny texts and making me feel like I learned something in graduate school when you called for help. The prayers and words of encouragement from my grandmother, Lois Cole, has remained a constant rock throughout my whole life journey. I am so grateful. I also dedicate this work to my grandfather, who would have been so proud to see me walk the stage in Missouri gold and black. I hope I inherited some of your wits.

I am so thankful for my extended village of aunts, uncles, and family friends, who often didn't know how a simple "proud of you" would spark a new fire in me to get this done. Thank you to my friends, near and far, for "hyping me up" when I didn't give myself enough credit. A special thank you to my cohort, for the nights of venting, marathons of movie watching, and just being there. No one else could understand this experience like you all.

Finally, but above all, I owe everything to God who has kept me and blessed me beyond all I could ask or imagine. I would be lost without Your grace.

ACKNOWLEDGEMENTS

This research was supported by two research grants from the Department of Psychological Sciences at the University of Missouri. A special thanks to my advisor, Dr. Nelson Cowan for helping me to develop as a creative, independent scholar. Thanks to my lab group as well as our lab manager, Bret Glass, for supporting the work of this dissertation. Additional thanks to my dissertation committee, Dr. Yuyan Luo, Dr. Jeffrey Johnson, and Dr. Elizabeth Kelley, for the invaluable feedback and kindness that helped me to complete the present work.

Table of Contents

LIST OF ILLUSTRATIONS

LIST OF TABLES

Abstract

In a previous study (here termed Study 1), we explored the effects of working memory loads on young children's recall of passive voice structures. We found that participants (n = 36) were more likely to use the passive voice in recall responses when holding an unrelated memory load but were more likely to use the active voice when not under load. In the two new studies presented here (Studies 2 and 3), I extend these findings to explore how working memory loads impact not only children's recall but also their construction of original passive voice structures. In the first new study, I used a similar method to Adams & Cowan (2021) but gave 4- and 5-year-old participants (n = 38) more instruction and practice recalling the passive. Responses were categorized as either passive or active voice sentences (a small subset of responses could not be categorized as either syntax). Participants were more likely than in the previous work to use the passive voice overall, but use was not significantly impacted by memory loads. Participants were more likely to use the active voice when there was no load, as in the previous study. In a second study, I explored if 4- and 5-year-old children (n = 36) could form their own passive voice constructions about animations depicting transitive actions. Participants used the passive voice in around 21% of all responses, and used the active voice for most of the remaining responses. Working memory loads did not cause participants to speak in the passive or active voice more often. Performance on the QUILS, which combined measures of vocabulary, syntax, and processing abilities, predicted participant's ability to use the passive voice, especially when under working memory load. Combined results demonstrate the difference in task demand between

recalling passive sentences versus constructing passive sentences and how working memory contributes to each.

Chapter 1. Introduction

A two-year-old Eloise visits the zoo with her dad and finds that some gates are too tall to look over without some help. From the ground, she raises her arms in the air and petitions her dad, “Up! Daddy, up!” She receives a lift from her dad that helps her to see over the tall gate to an enclosure of reclining lions. “Cat!” she exclaims as her dad chuckles. “Yes, those are big cats called *lions*,” he explains. Three years later, Eloise returns to the same exhibit at the zoo as a rambunctious kindergartner. Although she still needs a lift from dad, she now receives one by asking, “Daddy, can you pick me up to see?” and then remarks, “Those are big lions!” at sight of the big cats.

In the three years between Eloise’s zoo visits, she experienced a tremendous amount of learning and development that allowed her to expand her sentences and correctly label items. She begins to sound more like an adult and can gain even more information within conversation. What happened within Eloise’s brain and environment that caused such advances in her language skills from age 2 to age 5? In the two studies described here, I seek to answer this question by exploring how one specific brain-based resource, working memory, helps contribute to this development. I begin with a view that working memory can be considered the root of much of cognition (Lépine et al., 2005). As working memory capacity increases with age, all cognitive skills that depend on working memory to some extent also develop. Language processes such as production can be considered one branch on this tree of cognition. I seek to show how working memory is important for language production in the early childhood years, especially. It could be that, for Eloise, she has experienced considerable gains in working memory capacity between her two zoo visits. These gains might be part of the basis for her newly advanced speech.

Background

The field of working memory is relatively well-established with attempts as early as the 19th century to understand what seems to be the most immediate and brief form of conscious memory (James, 1890). In contemporary research, although the exact definition of working memory may vary slightly by view of theorists (Cowan, 2017), it is generally agreed upon to be the limited amount of information you can use to tackle ongoing cognitive tasks. For example, when trying to mentally calculate the change to give someone who has purchased a $3.55 item with a $10 bill, working memory helps to keep each step and number in mind, while executive processes carry out the arithmetic. Perhaps you're in a classroom of students and want each student to state their name. Working memory can also help keep track of who has already spoken and who still needs to speak by utilizing spatial cues (their seat in the classroom) and visual cues (their face or distinctive physical features). However, most people can testify to the feeling of incorrectly carrying out simple mental calculations or losing track of who has spoken in a group of people. These shortcomings shed light on the limited nature of working memory, which is one of the main characteristics that is said to set it apart from other forms of memory according to most researchers (Cowan, 2008).

This maximum amount of information that an individual can hold in mind is known as their *working memory capacity.* Although it is generally agreed upon that working memory capacity for the average individual sits around 3-4 items (Cowan, 2001, 2010), it can slightly differ from person to person. Even these slight differences can result in considerable disparities on various tasks and outcomes (Conway & Engle, 1996; Martini et al., 2020). Within an individual, working memory capacity also appears to be something that increases with age, until reaching stability in adulthood (Xu et al., 2018). This point has been debated, however, because some argue that other skills such as the ability to filter out irrelevant information, speed of processing, or the ability to rehearse items is what leads to better performance on working memory tasks. However, there is good evidence that there still remains some natural maturation of working memory that occurs within development, possibly due to biological changes like the development of critical brain regions (these debates will be discussed in the literature review section). For this reason, I find it important to analyze what changes in subsequent cognitive tasks are occurring because of this capacity growth.

It is possible to find out more about the limitations and utility of working memory by assessing what happens when individuals are required to complete dual tasks (Pashler, 1994), such as the ones described in the current set of studies. In these sorts of studies, participants are asked to perform at least one task known to require working memory. For example, the participant may have to remember a short list of words or digits, colors or locations of visual items, or a combination of both (Elsley & Parmentier, 2009). These tasks are often referred to as working memory *loads.* While this information is presumably held in mind by the participant, they are also asked to carry out a target task whose dependence on working memory is being explored. If performance on the target task is affected by performance on the working memory load, it can be concluded that working memory holds some influence on the target task. Sometimes, it is possible to find a double dual-task cost in that performance on both the working memory task and target task are negatively affected compared to a baseline (Doherty et al., 2019).

An individual's working memory capacity can serve as a moderator for how working memory loads affect their performance on the target tasks. In some cases, people with lower capacities can be more negatively impacted by loads (for example, Goldinger et al., 2003). In other instances, people with higher working memory capacity can be more affected by loads because of the heightened application of attention (Rosen & Engle, 1997). Either way, using dual tasks allow us to add experimental evidence to research exploring the use of working memory in cognitive tasks for all individuals. In the current studies, working memory loads are used to concurrently affect two different forms of language production. I suggest that the aspects of language production that become more difficult (as shown by less accurate use) or impossible (as shown by a complete lack of use) while under load are those that especially depend on working memory. It can then be presumed that with the growth of working memory, these aspects of language production will become easier to deal with, leading to more complex language use.

Research Questions

In the current set of studies, I explore the role of working memory in both the recall and construction of language, specifically passive voice sentences, by requiring children to hold unrelated memory loads. I assess how their recall and responses are affected by the working memory load, and vice versa. The following questions are of interest:

(1) How does working memory affect passive sentence recall?

(2) How do different types of loads (visual-spatial and auditory-verbal) affect passive sentence recall?

(3) Can children construct their own passive voice sentences about witnessed transitive actions?

(4) How do different types of loads (visual-spatial and auditory-verbal) affect passive sentence construction ?

(5) How do language skills and working memory capacity predict their ability to construct their own passive voice sentences?

(6) What are the differences in the task demand of recalling language versus producing original language and how does working memory contribute to both tasks?

By exploring these questions in the current studies, I wish to add to the overall understanding of how working memory contributes to young children's language production abilities. Specifically, I assess how working memory is important for the memory and construction of a syntax that may be less familiar or even novel to the 4- or 5-year-old ear. These findings may contribute to literature that seeks to understand what it is that allows children to acquire language and become masters of speech within the early childhood years.

Theoretical background

Working memory

The embedded processes model. The field of working memory is one with a long and rich history with varying theories about what working memory is, how it works, and when it matters (Adams et al., 2018, Logie et al., 2021). The present studies are primarily viewed through the lens of the embedded processes model of Cowan (1988, 1999; Appendix A). The embedded processes model proposes that all working memory processes are contained within long-term memory. Long-term memory is thought to be relatively unlimited in both time and space, given that memories can last for a lifetime and there is not a defined limit to the amount we can maintain. Within long-term memory is *activated long-term memory*, the portion of long-term memories which are given some heightened level of activation, although not yet full attention, making them more accessible for potential use. Activation is a way to describe higher amounts of neuronal-level activity distributed to representations held in mind (Cowan, 2019). Within activated long-term memory is the most-limited portion of memory and the highest activated information, referred to as the *focus of attention.* Given that this information is actively attended to, it is what the individual is consciously aware of at any moment. In order to manipulate information within the focus of attention and activated long-term memory, the model includes a set *of central executive processes* which allow the brain to carry out tasks on stored information. One of the hallmark characteristics of the model as a whole is the inseparable nature of the different subparts of the memory system. Although there are distinct processes, the information within working memory is proposed to be interdependent upon the other subparts of the system.

This embedded nature of Cowan's theory differs from other popular theories of working memory including that of Baddeley & Hitch (1974; Baddeley, 1998), which views working memory as a more siloed system. Their multi-component model posits that working memory is a system made up primarily of two specialized stores, one for auditory and verbal processing, the other for visual and spatial information. There is also an episodic buffer which is able to process information that combines both (or other) forms of information. In the multi-component model, long-term memory is separate from the working memory stores and central executive processes are able to move information in and out of the stores, as well as manipulate the information being held in stores.

Cowan's embedded processes model does not attempt to fully divide working memory according to modalities given the various number and combinations of modalities of information. However, the model recognizes that more similar information can interfere in activated long-term memory (Morey & Cowan, 2005).

Some theorists argue that short- and long-term memory should still be considered separate, distinct stores. To this point, Norris (2017) argues that the proposed activated long-term memory part of Cowan's model cannot explain how individuals remember repeated items in a sequence such as 5855 because only the items '5' and '8' would be activated in long-term memory. He goes on to suggest that there is no clear mechanism for denoting the serial position of these items in a list and a separate copy of information in short-term memory is needed to account for such patterns. Cowan (2019) explained that the memory system may allow for rapid learning such that serial information is organized in the focus of attention and the newly and rapidly learned information simultaneously becomes a part of activated long-term memory. This allows for the embedded nature of the model and does not require separate copies of information in separate short- and long-term stores.

Working memory limitations. Another characteristic of the embedded-processes model is its emphasis on the limits of working memory. Three kinds of limitations to what we are able to hold in working memory are considered: number of items, temporal decay, and stimulus interference. Memory researchers have attempted to discover the upper limits of what we can hold in mind for decades, with George Miller's 'magical 7 ± 2' emerging as the most popular descriptor after his seminal paper discussing the ever-present integer (Miller, 1956). Short-term memory studies asking participants to recall items in a list have found that there is often forgetting after people have reached about 5-9 items, seeming to confirm Miller's somewhat playful theory. However, later studies would show that when certain mnemonic strategies are controlled for, such as verbal rehearsal, item chunking, or grouping, a lower limit would be revealed. Cowan (2001) showed that in studies that use paradigms where the participant is no longer able to rely on these strategies, capacity limits appear to consistently fall around 4 ± 1 items for the average person. For example, in one measure of working memory, the running span, participants listen or view a running list of items that has an unpredictable end and are asked to state the last few items they can recall. Because participants are not able to tell when the list will end, they must constantly keep the last few numbers in mind and continue to update until given a cue to stop. Verbal rehearsal becomes virtually impossible, especially if the list is presented at rapid speed. Now, a purer capacity limit is revealed and given the task demands, we are better able to characterize the limits of *attention*, more specifically (Bunting et al, 2006).

While people may overestimate the number of items they can hold in mind at once (Cowan et al., 2016), most people including children, understand that our memory for information in the moment is limited by time. Whether we are trying to remember our new neighbor's name or the next task on a mental to-do list, it does not take long before the information can mysteriously slip from the mind. The embedded processes model recognizes that once information is no longer in the focus of attention, and therefore no longer protected by attention-based refreshing mechanisms (Vergauwe & Cowan, 2005), it becomes vulnerable to temporal decay. Researchers have debated the nature of this temporal decay, with some arguing that items just become less distinct from other items with time due to interference (Lewandowsky et al., 2008). Still others argue that this alone cannot explain the role of time-based decay, and that absolute temporal decay must also be considered. Ricker et al. (2014) was able to show through a series of experiments that manipulating distinctiveness of to-be-remembered items by varying the inter-item intervals was not enough to eliminate time-based forgetting. Even when participants are able to rehearse between items in a list, forgetting still occurs as a function of time. In order to prevent item decay, the individual must refresh or re-attend to items through recall or repetition. Once the item has been supported enough through attention-based mechanisms, a process known as working memory consolidation (Cotton & Ricker, 2021), it can reach a level of permanence that allows it to last for much longer periods of time without being attended to and is considered to be a part of long-term memory (e.g., your memory of a childhood birthday party).

Information can also be susceptible to interference when not in the focus of attention. While some theories propose interference as the main source of forgetting (Underwood, 1957; Lewandowsky et al., 2008), the embedded processes framework recognizes it as one process that can occur among others. As noted, information is especially vulnerable to interference within activated long-term memory (Cowan et al, 2014). Items that overlap in features (e.g., phonological, visual, tactile) can be confused within the working memory system and retrieved at the incorrect time, which appears as forgetting on tests of memory. Of the different types of limitations (or obstacles, in this case) that information must face in memory, interference seems the most difficult to prevent. Individual differences in working memory capacity and attention control can sometimes predict susceptibility to interference (Kane & Engle, 2000).

In sum, the field of working memory contains a well-established and still expanding body of literature that explains how this resource drives many of our daily cognitive tasks. The embedded-processes framework allows us to examine how *attention* more specifically is limited, and how it relates to these daily tasks. One of those such tasks is language production, which I turn to next.

Language production

The ability to produce language that other individuals can comprehend is arguably one of the most unique characteristics of human development and interaction. Though it is a universal trait, language production is a complex task that requires a significant amount of learning and effort, across the lifespan. Like the field of working memory, linguists have attempted to theorize how the language production process works for decades (c.f. Pickering et al., 2013). Given the current set of studies' emphases on grammatical production and speech errors, I use the Bock & Levelt (1994) model of language production as a framework for interpreting the present results.

Bock & Levelt coined their theory of the language production process *grammatical encoding* (re-illustrated in Appendix A). In it, the speaker must begin with an intended message which is prepared to undergo the grammatical encoding portion of the process, next. Within the grammatical encoding portion are two types of processing: first *functional*, then *positional*. Functional processing includes the steps of *lexical* *selection*, in which the correct words must be selected from the speaker's vocabulary to match the intended message. Also included is *function assignment*, which assigns grammatical roles or syntactic functions to each of the lexical items in the forming sentence. Next, within positional processing, the message undergoes *constituent assembly* which creates ordered word slots for the selected lexical items and also *inflection* which adds small elements to lexical items to indicate number or tense (e.g., -ed to represent the past). Once these processes are complete, the message enters the stage of phonological encoding and is sent off to output systems to be spoken, written, or hand-signed.

The model of Bock & Levelt (1994) is especially useful for understanding why speakers produce errors. The authors suggest that errors can happen at any level of the model, and become evident in the nature of the error. For example, if a person says "Let's get on the car" instead of "Let's get on the bus," it can be presumed that the error occurred during the lexical selection part of the process, likely due to semantical overlap between *car* and *bus*. If the error occurs at the level of function assignment, a sentence such as, "We are going to the milk to pick up some store," may emerge. Bock and Levelt propose that all these errors can occur because of feature overlap. When words or clauses that are similar in feature (phonological, semantic, etc.) to the intended message are activated, it is easier for the incorrect word or clause to be produced. This view connects well with the embedded processes model framework because of the activated portion of long-term memory. Lexical items that are similar to those that are heard or thought of become activated to be potentially selected into the focus of attention and used for the present language task. However, if there is some level of distraction or if working memory is already taxed, central executive processes may select a similar but incorrect word to produce. Likewise, distractions or loads can cause words to be placed into the incorrect word slots, causing function assignment or constituent assembly errors. The different types of errors that occur in the present set of studies can be easily characterized by a disruption at some level of the Bock & Levelt model of language production.

The present studies

The present studies aim to explore how working memory contributes to the ability to produce language, specifically in children. Early childhood is an opportune period to study the interaction of these mechanisms because of the relevant language abilities that emerge during childhood. With this work, I hope to better understand how working memory contributes to language development, one of the pillars of general development. In the following, I begin by reviewing the relevant literature related to working memory and language development over childhood. I also highlight the purpose of the use of the passive voice in the current methods. Then, I will briefly recap a previous study that has formed the foundation for the present pair of new studies. The two new studies to be described will explore what occurs when children are given increased motivation to use the passive voice during sentence recall and what occurs when children are asked to formulate their own passive voice sentences. Using these two methods, I will be able to ask how working memory contributes to different language tasks. I will conclude with a general discussion about what the studies combine to illustrate about the relationship between working memory and language production in young children, the implications for everyday use, and future directions for continuing to address unanswered questions about these issues.

Chapter 2. Review of Literature on Development of Working Memory and Language Production

Working memory development

Like many skills children acquire with age, working memory undergoes an impressive amount development over the childhood years. Gathercole et al. (2004) tested children ranging from ages 4 to 15 on batteries of working memory, specifying tests that required verbal, visual, or complex (mixed) memory. Results revealed consistent linear gains in performance, even into adolescence on all three types of measures. What causes these increases in performance accuracy on working memory measures? Theorists have debated on the underlying causes, with some arguing for gains in knowledge and skills and others arguing for a passive growth of capacity that occurs as neural structures develop.

One common explanation for positive changes in working memory test performance are gains in knowledge (Case et al, 1982, Baxter et al., 1996). As children begin to learn more words and concepts, keeping these items in memory become less effortful. For example, memorizing a list of fruits such as banana, orange, pear, and grape may be easy for anyone familiar with the concept and names for these foods. Recalling each fruit should be fairly easy given the prior knowledge and even facilitate grouping the items (e.g., forming pairs such as banana-orange and pear-grape). However, if asked to remember names of professional basketball players such as James, Westbrook, Harden, and Lillard, a knowledge divide may become more evident. Avid sports fans should have no problem recalling this list and may even imagine a court with the players arranged according to their positions. Someone who has no knowledge of professional basketball might have to attempt to memorize the names almost as random words, which makes mnemonic strategies more difficult. Thus, theories that propose that working memory gains are largely due to knowledge increases do hold credibility. Halford et al. (2007) note how knowledge helps us to chunk information into meaningful items, which allows for more capacity to store additional information.

Despite the necessity for and importance of knowledge gains, they cannot wholly explain why children improve on working memory measures with age. Even in adults, knowledge only tells part of the story as Hambrick and Engle (2002) demonstrated by showing that at high levels of expertise, working memory capacity is still able to predict individual difference in retainment of information. Cowan et al. (2015) also explored the role of knowledge by requiring child participants (aged 7 – 13 years) and college-aged adults to remember groups of English letters and unfamiliar characters. While memory for English letters was certainly better, when scores on memory for both types of items were standardized, nearly identical developmental trajectories emerged. Memory recall for both familiar and unfamiliar items still increased with age at similar rates, suggesting that a capacity-based growth mechanism was still contributing to gains in performance. Although knowledge plays a role, it cannot be considered as the sole mechanism behind the increase in working memory performance with age.

Another explanation for the increases in working memory performance is the development of the ability to filter out irrelevant information and distractors. In everyday life, focusing attention on one thing typically means having to filter out irrelevant stimuli from multiple other sources (e.g., other voices in a room or other words on a page). In lab settings, working memory researchers can explore how individuals filter out irrelevant information by using paradigms that present multiple items simultaneously. In one such study, Vogel et al. (2005) were able to show that participants with lower working memory spans had a more difficult time filtering out irrelevant information, which likely led to results demonstrating low spans. Part of the task for participants in working memory studies is to be able to only remember items that they will be tested on, otherwise interference can occur and cause error in response. This process is known as *selective attention* and is one that has certainly been shown to improve with age (Rueda & Posner, 2013; Booth et al., 2003). However, like gains in knowledge, changes in attention filtering abilities do not fully explain why working memory capacity appears to improve with age. Cowan et al. (2010) tested the role of filtering by requiring children ages 7 to 8 and 12 to 13, as well as college-aged adults to remember arrays of colored shapes. In some conditions, participants were instructed to only pay attention to specific shapes and ignore others. Results showed that the youngest children performed the same as older groups at allocating attention to the correct shape set (especially for lower set sizes). However, younger participants remembered significantly fewer items than the older aged participants, suggesting that a difference in capacity was still present, even when attention allocation was similar. A growth of capacity hypothesis still seems to be necessary when attention filtering abilities are controlled for.

Lastly, and of special importance for work examining language development, some researchers argue that the development of verbal rehearsal strategies help promote better performance on working memory measures. An effective strategy to remembering information that can be verbally encoded is rehearsing (or simply repeating) the information until it is used or becomes more permanent in memory. Previous work has suggested that children likely do not rehearse information on their own until around 6 years of age (Flavell et al., 1966, also see Elliott et al., 2021). Verbal rehearsal is one of the highlights of Baddeley's model of working memory, in which the phonological loop depends on rehearsal to keep auditory information in store (Baddeley, 2003). If this is a strategy with a relatively late onset and continued improvement with age, it should make sense that rehearsal is a driving force behind working memory development. To test this contribution, Cowan et al. (2011) asked participants to remember an array of visual items while either naming the colors out loud, remaining silent, or engaging in articulatory suppression during the task. Although articulatory suppression did cause poorer performance, there was no difference in its negative impact between children and adults. Overall, it is likely that gaining rehearsal abilities will help the developing child with remembering information, but these abilities still cannot fully explain why working memory grows with age. The answer seems to still lie in a growth-of-capacity mechanism.

If working memory capacity is what is naturally increasing with age, then what allows for this growth? Much of the cause can be attributed to maturational changes that occur with age in relevant brain structures. In one study, children between the ages of 9 and 18 completed a visuospatial working memory task while receiving an fMRI scan (Klingberg et al., 2002). Older children appeared to show more activation in areas such as the superior frontal sulcus and intraparietal sulcus. Activation in both these areas correlated with higher working memory capacity. The authors suggest that the maturational changes may be due to processes such as myelination and synaptic pruning. Relatedly, Thomason et al. (2009) also required participants (aged 7 – 12) and adults (aged 20 – 29) to complete a visual working memory test while undergoing an fMRI scan. With increasing set size, children made more errors in responses and also did not exhibit the same amount of increasing activation as adults. This suggests that although activation mechanisms are similar in children and adults, they are not yet fully developed, leading to increases in error and therefore smaller capacity limits. Differences in structural and functional brain imaging results can even be used to predict working memory capacity later in life, serving as testament to the importance of natural development for the growth of working memory (Ullman et al., 2014).

Given the clear literature demonstrating the growth of working memory with age, it appears to be a critical resource for cognitive development. Yet many studies of working memory do not attempt to generalize the role of working memory beyond that of laboratory tasks which often do not closely reflect tasks in life. Why does the development of working memory matter? For one, working memory has been shown to be a reliable predictor for many skills required for school and later life achievement. For example, one of the most common measures for school achievement, reading skills, can be predicted using working memory paradigms. The most popular reading-related working memory test is the reading span test in which participants read sets of unrelated sentences and attempt to recall the last word of each sentence at the end of each set (Daneman & Carpenter, 1980; Friedman & Miyake, 2004). Participants who can recall more words also score higher on tests of reading comprehension. The ability to store these words in mind while also processing each sentence through reading requires working memory and the words that can be remembered are one way to measure participants' working memory capacity. Working memory has also been shown to be a predictor of math skills and reasoning (Bull & Scerif, 2001; Simms et al., 2018).

The present works seeks to add to past literature on the relationship between working memory capacity and the ability to communicate (Määttä et al., 2014). Language production fits in as one of the many cognitive tasks that rely on working memory for successful funcitoning. I propose that as working memory capacity increases, so too does the ability to produce language, a critical skill which will be reviewed next.

Development of language production

The typically developing child will most likely utter their first distinct word by the first year of life (Bloom, 1976/2013). Before then, much of language production comes in the form of babbling or other vocal sounds, that often reflect similar sounds and sequences as meaningful speech and predict later language skills (Oller et al., 1976; Vanormelingen et al., 2020). Around 18 – 20 months, children start to combine words to form short sentence-like structures, originally referred to as *telegraphic speech* (Gerken et al., 1990). There is even evidence that these short utterances can hold more meaning than initially perceived, as told by previous work on *semantic relations* (Golinkoff, 1981). These two-word combinations can reflect possession, intended actions on an object, or actions observed by other agents (Bloom et al., 1980). Over the next few years, children will continue to gain words to add to their rapidly expanding lexicon through processes like *fast-mapping* (Carey & Bartlett, 1978, Alt et al., 2004). Though they will face obstacles to overcome such as overregularization (Marcus et al., 1992) and overgeneralization (Ambridge et al., 2013) of grammatical elements and nouns, language production abilities start to become mostly adult-like by the beginning of the school years.

The development of language abilities over the first few years of life is undeniable, but much like the development of working memory, theorists have differed in their view of the *cause* of language development. Most hypotheses about language development can fall into one of three overarching theory camps: nativist, interactionist, and cognitive. The nativist view of language development emphasizes that the unique skill of human language is one that is hard-wired into our biology. Much of this view is led by theorists such as Chomsky (1995) who proposed that children will naturally learn language being spoken around them without much effort, which must be some evidence for an innate language learning mechanism. He proposed the *language acquisition device* to especially explain children's ability to rapidly learn grammar structures, even given the complexity and breadth of rules that must be acquired. Moreover, he proposed that children are born equipped with a set of universal grammar rules that are common to all languages that allow them to build upon and expand as they grow. More contemporary nativist theorists argue that evidence such as young infants' knowledge about syntax still support the idea of universal grammar (Lidz et al., 2003; Lidz & Gleitman, 2004).

Many researchers who subscribe to the second theory camp, the interactionist view, might agree that children do seem to be born with language acquisition abilities. However, the process of acquiring language is deeply dependent upon experience including social interactions, environment, and direct learning from input. There is a substantial body of literature demonstrating how language input is a consistent predictor of language development (Huttenlocher, 1998). Parental vocabulary size and syntax abilities can predict a child's subsequent knowledge and ability (Hoff-Ginsberg, 1990; Pan et al., 2005). Socioeconomic status has been a consistent predictor of child language ability, with children from lower income backgrounds often falling behind in measures of complex syntax skill (Pace et al., 2016). Even non-parental caretakers such as teachers can impact children's abilities (Farrow et al., 2020). Overall, what children hear in their environments hold undeniable importance for the way their language skills develop.

The third camp of theory is not as well established as the former two, but describes views that attribute much of language development to the growth of cognitive capacities. These theorists often view language development through an information-processing lens, highlighting that as the brain develops and cognitive resources mature, skills like language are able to flourish. As resources such as speed of processing (Peter et al., 2019), perception (Casserly & Pisoni, 2010), and attention (de Diego-Balaguer, 2016) develop, so too can the ability to comprehend and produce language.

As with most theoretical discussions, the answer to the question of which mechanism drives language development likely lies within a mixture of each of the three views presented here. Understanding how and why young speakers' skills improve with age is probably best done by considering what abilities they are already equipped with at birth, how their unique environments support and build on these abilities, and how their cognitive capacities support this building. The present work aims to add more evidence from the cognitive perspective by emphasizing the importance of working memory for language development. I propose that working memory is perhaps one of the most central and necessary cognitive resources for language development, and the growth of the former allows for the growth of the latter. Of course, much work needs to be done to be able to draw causal connections and this need has led to the present set of studies.

This work focuses on how working memory contributes to the ability to perform two related, but different language tasks: reproducing previously heard sentences and producing new, original sentences. In both cases, an uncommon syntax (for young speakers) is used to understand how these processes function during early acquisition stages. Next, I discuss existing literature about the use of working memory for these two language tasks.

Working memory for different language production tasks

Reproducing language

Interactionists would likely agree that some part of language acquisition can be achieved through simple repetition of words and phrases heard in the child's environment. A young listener who hears his mom exclaim she is "exhausted" after every long day of work may also complain of being exhausted, even if the context is not yet fitting. Repetition and imitation serve as tools for gathering new language elements and practicing them in conversation with others. Some even propose that the ability to repeat syllables and words is a fundamental aspect of working memory. Nonword repetition tests require children to listen to lists of nonsense syllables increasing in length and repeat them verbatim. Researchers who use nonword repetition measures propose that not only can they be used as measures of working memory capacity, but as a predictor for language acquisition skills (Gathercole et al., 1994; Gathercole, 2006).

Beyond syllables, a child may also be asked to repeat whole sentences to assess mastery of language and to reveal potential difficulties with language processing

(Polišenská et al., 2015). Some studies have demonstrated the importance of working memory for the ability to recall full sentences. Alloway and Gathercole (2005) tested 4- and 5-year-old children on multiple working memory measures (specified as phonological memory tests) and split participants into low and high span groups. Participants were then asked to recall 20 sentences varying in length and grammatical structure. Responses were scored for accuracy of recall and occurrence of production errors. Overall, participants with lower phonological memory scores were more likely to produce errors and more negatively impacted by serial order effects. This study serves as a good example of the importance of working memory for even the immediate recall of short sentences. However, some scholars suggest caution before making these claims, as much of sentence repetition appears to rely on language skills themselves, as opposed to working memory capacity (Klem et al., 2015). It is worth noting that the vast majority (if not all) of the literature assessing the relationship between children's ability to recall sentences and their working memory abilities have been correlational in nature. While certainly informative, there are limitations to the claims that can be made for or against any connections given that working memory measures are often cyclical (i.e., you need proficient language skills to score well on tests of verbal working memory). The current work seeks to add to this body of literature by assessing how *constraining* working memory affects children's ability to recall sentences after a delay. By essentially tying up part of every child's working memory, no matter their capacity, it is possible to assess how working memory is useful for language tasks happening in the moment.

Producing original language

In addition to repeating words or phrases, children are also tasked with likely the most difficult but important language task there is—devising new sentences that accurately portray their thoughts and ideas or describe events in their environment. This requires using their acquired vocabulary and syntactic rule knowledge to form sentences that portray their intended message and can be understood by their conversational partner. Some studies have explored children's ability to do so by eliciting production. For example, in the "cloze procedure", participants are given a partial sentence with a missing word and are asked to complete it (e.g., "Chickens cackle and ducks ______."; Taylor, 1953). This task has been used to elicit parts of sentences from children when they may not normally form the sentences on their own. Bellon-Harn et al. (2004) describe the use of the procedure for interventional work with children who have language impairments. Children, 5.5 – 6 years of age, with language or phonological impairments, engaged in a storybook reading task with an experimenter who used a combination of conversational methods including the cloze procedure (e.g., "The cow jumped in the mud and –"). This method in combination with other types of sentence completion and expansions helped to increase the complexity of sentences children used in their own speech. Other researchers interested in typical development have also used sentence completion tasks to test children's understanding of specific syntactic or inflective rules (Berko, 1958).

Possibly due to the complex and abstract nature of this ability, less work has been done to explore how working memory is related to children's ability to produce original sentences. One study to do so measured preschool children's phonological memory skills and also analyzed spontaneous speech during play (Adams & Gathercole, 1995). Children with better memory scores were more likely to produce longer and more complex spontaneous utterances. Blake et al. (1994) asked children aged 2 to 5 to repeat list of animal names (a test of verbal working memory) and also imitate sentences. Recordings of their spontaneous speech revealed that children who scored higher on the recall and imitation tasks were more likely to produce lengthier sentences in their own speech. The authors suggested that the results revealed the constraints of memory on children's speech production capabilities, one of the primary goals of the studies described here.

Though there is extant evidence on the connection between working memory and spontaneous speech, not much work has examined the role of working memory in children's ability to describe specific events in the environment. Part of the challenge with this angle of research is being able to discern when differences in performance are due to variance in working memory or variance in language skills. Being able to describe events may require working memory, but it also requires the words and syntactic structures to be able to do so. In order to begin to answer this question, the present work uses an experimental approach to constrain all participants' working memory as they describe animations. Though children may approach the task with varying lexicons and syntactic skills (both of which will be measured in separate auxiliary tasks), the use of working memory loads will allow us to view how constraints affect *all* speakers, regardless of language abilities. I will then be able to assess how much of a contribution working memory makes based on causal effects and how much of a contribution vocabulary, syntax, and acquisition skills make based on correlational relationships. In both of the new studies described in the present work, the focus is on children's ability to recall or produce the passive voice in particular, due to its relatively infrequent use in child speech. Next, I review the nature of the passive voice as well as its usefulness for answering questions about how working memory contributes to the ability to use complex and unfamiliar language.

The passive voice

Transitive-action sentences are those in which an agent(s) acts upon a patient(s) in some way. These sentences can be stated to either emphasize the agent or the patient of the action, by using either the active or passive voice. For most native-English speakers, the active voice is considerably more common in natural conversation, though the passive voice tends to be used more often in text (Dąbrowska & Street, 2006). Even early in development, toddlers show biases for the active voice. Abbott-Smith et al. (2017) used eye-tracking and forced-choice paradigms to assess children's bias for interpreting the first noun phrase of passive and active voice sentences. Both 2- and 3-year-old showed a bias for interpreting the first noun-phrase as the agent of the sentence, but both were able to resolve their interpretation after hearing the rest of the sentence, if it was stated in the passive voices. However, 2-year-olds only showed evidence of this resolution in the eye-tracking condition rather than the forced-choice condition, while 3-year-olds showed evidence for resolution in both cases. This not only reveals evidence for how early active-voice biases can form, but also about how task demand can overshadow what children are capable of.

There is mixed evidence on exactly when young children are able to comprehend and then use the passive voice, with most citing limited comprehension and production in the 4–6-year-old range (Baldie, 19761, Bidgood et al., 2020). Younger children have been shown to be able to sometimes use the passive voice, but especially if the sentence is elicited. For example, Brooks and Tomasello (1999) taught 2- and 3-year-old children novel verbs using either passive or active voice sentences to describe interacting toys. In a separate session, participants were asked again about the toy with a sentence that either elicited the passive voice (*What happened to the duck?)* or active voice (*What did the car do?*). The vast majority (90%) of participants who heard the novel verb in a passive voice construction and were then asked a passive-eliciting question were more likely to use the passive voice in response. Brooks and Tomasello argue that these results reveal that one of the primary reasons for this relative delay in comprehension and production is the lack of use of the structure in child-directed speech (Gordon & Chafetz, 1990).

Others argue that the level of difficulty in understanding or producing the passive voice can also vary specifically by the type of verb used or the use of the *by*-phrase (Hirsch & Wexler, 2006; Fox & Grodzinsky, 1998). Whether these delays are due to a lack of input or the unique structure of the passive voice sentence (or a combination of these factors), these challenges make the passive voice a good candidate for studying novel sentence construction in 4- and 5-year-olds. This age group is also in a developmental sweet spot when most are beginning to produce adult-like structures (better than their 3-year-old counterparts), but still underdeveloped in some language tasks (e.g., avoiding overregularization) than their older peers. Here, I assess how working memory contributes to 4- and 5-year-olds ability to recall passive voice sentences as well as produce their own passive voice sentences. Using these methods, I hope to better understand how working memory is useful for the acquisition of new language structures in early childhood.

Chapter 3. Study 1: Recalling the Passive Voice Under Concurrent Load

In Adams & Cowan (2021), we sought to find how working memory loads affect 4- and 5-year-olds' ability to recall sentences in the passive voice (I will refer to this as Study 1 in the remaining chapters). At the beginning of the study, participants (n = 36) were introduced to the passive voice, labeled as a "special way" of describing pictures. They were instructed to attempt to use the same special way in all their responses throughout the experiment. Then, in three experimental blocks, participants were shown images depicting a potential patient and agent and heard the experimenter describe it using a passive voice sentence (e.g., *the balloon was popped by the giraffe*). Half of sentences were reversible in that the agent and patient of the sentence could be switched and still maintain roughly the same level of logic (e.g., *the horse was pushed by the cow* could also be *the cow was pushed by the horse*.). The remaining half were irreversible so that switching the patient and agent of the sentence would affect the logical likelihood (e.g., *the flower was watered by the girl* could not be reasonably switched to *the girl was watered by the flower*.).

In two of the three blocks, participants were asked to hold an additional working memory load: visual-spatial or auditory-verbal. Visual-spatial loads were 4 x 4 grids of empty boxes. For each trial, 2 or 3 gold tokens (or coins) would each be dispersed inside 2 or 3 of the boxes. Participants were asked to remember which boxes had tokens to draw them after a short delay on a dry-erase response sheet. Auditory-verbal loads required participants to listen to the experimenter state 2 or 3 digits in a list at approximately 1 digit/second. Participants were asked to attempt to remember the digits to be restated after a short delay.

During experimental blocks that included these two types of loads, participants were asked to recall the description of each image during the short delay between probe and recall of the external load. In the third block, there was no additional load and participants simply recalled the passive voice descriptions of images. The order of the blocks was counterbalanced to ensure changes in recall performance could not be explained by practice effects, alone. Participants used the passive voice ($M = 0.65$, $SE = 0.05$) more than the active voice ($M = 0.26$, $SE = 0.04$) in their responses. We also found that the rates at which they used the passive or active voice varied by the type of load participants were required to hold. Participants used passive voice the most when they were under visual-spatial load ($M = 0.73$, $SE = 0.05$), especially compared to when there was no load ($M = 0.55$, $SE = 0.05$). Participants used the active voice the most when there was no load ($M = 0.38$, $SE = 0.05$), compared to both visual-spatial ($M = 0.18$, $SE = 0.05$) and auditory-verbal loads ($M = 0.23$, $SE = 0.05$). These results were the opposite of our expectations, which predicted that passive voice use, being the more unfamiliar and challenging syntax, would require more working memory and thus be easiest to use with no external load. Instead, we found that participants were more likely to paraphrase the sentences to the more common active voice when there was no load.

Our results led us to new hypotheses that perhaps attempting to simply repeat a sentence as you heard it, even if the syntax is less familiar, is easier than paraphrasing the sentence to an alternative syntax while maintaining semantics. The latter task requires comprehending, transforming, and producing a sentence correctly, while the former task only requires some level of mimicry. To add on to this possibility, we found that participants were more likely to produce errors in speech when trying to use the passive voice. This may suggest that there was not complete comprehension, but only rote memory of the sentence itself, which likely requires only shallow processing and is more susceptible to errors. Moreover, results showed that participants were more likely to produce active voice sentences that were semantically correct when the sentences were irreversible, suggesting that the range in possible ways to describe the image in combination with the memory load affected children's choice of syntax.

The unexpected outcomes of Study 1 led us to suspect that, when there was no additional load, participants were more concerned with portraying the correct meaning of the intended sentence, rather than using the target syntax. In Study 2, I attempt to further encourage children to speak in the passive voice to be able to better assess passive voice recall and error-rates, which would help to further understand whether children comprehend the sentences they state.

Chapter 4. Study 2: Emphasizing the Importance of Syntax

In Study 1, we found that children were capable of using the same passive voice construction in their recall of sentences. However, when they were not under load, they were more likely to transform the passive voice image description to the active voice, as compared to when they were required to hold an additional load. Among other conclusions, we hypothesized that it could be the case that children were especially concerned with producing the correct *meaning* of the sentence, rather than the target syntax. When faced with an image in which the child understands the action and knows the name of the patient and agent, it may be that the way to ensure that semantics are correctly communicated is to use a syntax that is more familiar and possibly easier to understand.

For the next study, I explore what might take place if there is an emphasis on the importance of attempting to use the target syntax, passive voice. The method for Study 2 is a replication of Study 1 with a few key changes: (1) more training on the passive voice with a memory and feedback component, (2) repetition of each probe sentence, (3) slight changes on the load stimuli to create easier, but hopefully more motivating load trials.

First, in order to provide more training on the passive voice, still referred to as the *special way* of speaking, the instruction and practice block was extended. Now, based on responses from Study 1, the experimenter explained to children what alternative sentences they may be tempted to use, but explained that these are not the special way (i.e., the active voice version of the sentence or a passive voice version of the sentence which incorrectly switches the patient and agent). Each session began by presenting four practice trials in which the participant saw each image and heard a passive voice sentence and were asked to repeat each one before moving to the next. Shortly after, children were told they would begin the first memory component by recalling the descriptions of the images they just heard. If they did not understand the premise of using the special way like the stimuli, the experimenter could provide feedback to help further explain. The vast majority of children immediately understood to recall the same sentence that was previously heard. We asked children to try to use the special way throughout the entirety of the experiment, if possible.

In order to further boost children's likelihood of being able to recall the passive voice image descriptions, I modified the experiment from Study 1 to now require participants to immediately repeat each probe description as they heard them before moving to the next. This act of motivated repetition could encourage children to engage in verbal rehearsal, a strategy that does not seem to be fully developed in this age group of 4- to 5-year-olds (Flavell et al., 1966, Elliot et al., 2021).

Another small change made to the experiment stimuli was to have participants repeat every verbal load (still digits) immediately after hearing them. For the visual-spatial loads, I reduced the grid from a 4 x 4 array to a 3 x 3 array. The aim was that this repetition on the auditory-verbal loads and reduction of possible locations for the visual-spatial loads would make it easier to encode and recall load stimuli, thus increasing the effect of loads on the language task. Another overarching change to this study was that it was all conducted online, via video conference call, to be described in the next subsection.

Online Experimentation

Due to the global pandemic beginning in 2020, the entirety of our research practices were transitioned to online formats. It could go without saying that the participation experience for children changed considerably, the effects of which may continue to unfold in the coming years (cf. Sheskin et al., 2020, Lourenco & Tasimi, 2020). For the current study, the experiment was programmed and conducted using an online experiment software platform, Psytoolkit (Stoet, 2010, 2017). Each of the sentence stimuli remained the same and each of the audio descriptions and the digits for the verbal load were uploaded into the experiment software. Children could complete the study on a tablet, laptop, or computer in their own home, with the assistance of their parents. Depending on what was most convenient for the participant and parent, the experiment was either completed through their own internet browser, or through remote control of the experimenter's screen on Zoom, though the experiment appeared the same either way. Parents were asked via email before the experiment to refrain from providing too much help for their children, beyond pushing buttons and providing repetitive explanations, where necessary. The experimenter kept note of any major distractions (e.g., the child being interested in something else in the room for long periods of time, internet disruptions, etc.). The participants described here today were those that completed the study with minimal distractions.

Hypotheses

The modifications to the method had the potential to cause three major changes in results. First, given the greater emphasis and explanation of the passive voice compared to Study 1, I expected that participants would use the passive voice more often than the active voice and at higher rates than in the first study. Alternatively, if participants hold a strong enough bias towards using the active voice or producing sentences that match intended meaning (and therefore paraphrasing), these rates may not increase from Study 1.

Second, I expected working memory loads to be even more impactful on passive or active voice use than in Study 1 because of the increased effort to use the passive voice. If working memory loads cause more reliance on repetition or mimicry as suggested in Study 1, this reliance will increase even more given the emphasis on the use of the passive when children find it possible, thus causing working memory loads to appear more impactful by increasing passive voice use. Alternatively, working memory loads might be less impactful than in Study 1 because participants get more practice repeating the sentence and thus do not need as much working memory to recall.

Third, I expected participants to produce more errors given the greater emphasis on using the correct syntax rather than semantics. However, there is the possibility that participants will produce the same amount or even less errors because of the newly added requirement to repeat each sentence. These repetitions may serve as a rehearsal mechanism that increases the chances of saying the sentence correctly.

Overall, the results will help us to better understand the nature of why working memory loads caused Study 1 participants to rely on the passive voice more under load. It should also indicate if simple repetition is the resource children are using to do so.

Method

Participants

For this online study, I recruited participants primarily through Facebook advertising, as well as an electronic ad to parents with children in the local school district. Thirty-eight participants ranging from 4:0 to 5:11 (MA: 5:0) completed the study in full with minimal distractions. With a sample size of 38, and standard deviations around 0.25, mean differences of 0.11 should be detectable and sufficient enough for a conclusive Bayes factor (as calculated by the BayesFactor package in R; Morey et al., 2018). I re-address these criteria in selected analyses in order to assess the power of the current sample size. Of the completed studies, 28 participants were White, 3 were Black or African American, 3 were Asian, 4 were more than one race. All children had normal hearing and normal or corrected-to-normal vision. Two participants were reported as having a first language other than English first, but reported fluent English skills by the time of testing. Half of participants were female.

Design

As in Study 1, there was an additional memory load task for two of the three experimental blocks: one block required participants to hold a visual-spatial load and the other, an auditory-verbal load. The third block contained no load.

The order of the experimental blocks were counterbalanced so that participants could receive one of six different block orders. The 24 images used during trials were grouped into 3 groups of 8, so that half of images were reversible and the other half were irreversible in each group. Sentences which were reversible contained an agent and patient that could be switched and still contain the same level of logic (e.g., *the cat was* *chased by the dog* could also be *the dog was chased by the cat*.) Sentences which were irreversible included an agent and a patient that could not be switched and still maintain logical reasoning (e.g., *the balloon was popped by the giraffe* makes sense, but not *the giraffe was popped by the balloon.*) This manipulation was included to assess how the nature of passive voice sentences can further affect ability to accurately recall and rephrase.

These three groups of images were counterbalanced so that they could appear in three different orders, creating a possibility of 18 order combinations of block and image group. Beyond this counterbalancing, the order of the image descriptions, trials, and loads were completely randomized by the experiment software making it highly unlikely that any two participants received the exact same order of stimuli and tasks. Due to the counterbalancing constraints of the software, images were anchored to a specific load size, if applicable. For example, if an image appeared with a *visual* load size of 2 for one participant, it could appear with a *verbal* load size of 2 for another participant. However, it could also appear under no load for yet another participant. I kept this consistent to ensure that some participants were not having to recall images under *heavier* loads than other participants, although they could be under different types of load.

Stimuli

As mentioned, for this current study, all stimuli and tasks were programmed into an online study. Identical images and sentence stimuli were maintained from Study 1. For this study, some images were adjusted so that the action of the sentence could not be more easily construed than other images and still required some level of memory. For example, if the target sentence were "The duck was chased by the goat," and the image portrayed a duck and a goat both facing to the right (with the goat on the left), it might be easier to ascertain the action of chasing. In this case, the image was adjusted to display the duck and goat facing each other. The images remained the same otherwise, and were displayed on light gray background.

Visual-spatial loads were now displayed as a 3 x 3 grid of white boxes on a black background. Participants still saw 2 or 3 gold tokens displayed in the boxes for each memory load. Auditory-verbal loads were recordings of digits spoken by a female voice at the rate of approximately 1 s/digit. The auditory-verbal load stimuli were created from digits 1 – 9 and were not allowed to repeat in the same list (e.g., 585 would not be used as a digit list). For both types of load, half of trials were a set size of 2 items and the other half were a set size of 3 items.

Procedure

Each participant's parent or guardian were sent a link to the online experiment before the meeting time. At the beginning of the Zoom session, parents were instructed to share their screen and navigate to the experiment link. For some sessions, technical difficulties made screen sharing too difficult or impossible. In these cases, the experimenter would begin the experiment and share the screen with the parent and participant. Remote control of the experimenter's screen would be provided to the parent so that they could complete the study through the Zoom interface. Parents began by providing consent and completing a demographic form. Then, the experimenter obtained verbal assent from the child after explaining the purpose of the study and their rights as a participant. The experiment itself typically began about 5-10 minutes after setup and form completion.

The experiment began by familiarizing the parent and participant with the buttons (e.g., next and back) that would appear and an explanation about how the experimenter would be leading them through the "game". The parent or participant (depending on the child's comfort or interest with clicking buttons) clicked a Next button to proceed throughout the introductory screens, as the experimenter read the contents.

Instructions began by telling the participant that they would view various pictures, listened to what happened in them, and play a couple memory games. They were then told that the pictures would be described in a special way and that they would have to try to speak in the same way for the duration of the experiment. Participants were first provided with one example that depicted ice cream and a sun. They were told, "If I want to use my special way of talking to describe this picture, I say, 'The ice cream was melted by the sun.'" Then, participants were shown two ways of describing the picture that were *not* the special way (i.e., "If I said, "The sun melted the ice cream," that makes sense, but it is not my special way!" and "If I said, "The sun is melted by the ice cream," that's a bit too silly! It almost sounds like my special way…"). Participants were told to be careful and try not to use these other ways of describing images.

Then, participants were told they would listen to the audio descriptions of 4 more images as they appeared one at a time. They were instructed to repeat the description and press a red button which said [I said it! Next →] after they repeated the sentence. After repeating every description and a short delay, participants were instructed to try to recall those same image descriptions in the special way, which served as the first practice memory test. One by one, the participant viewed each image and stated the description they could remember. They were encouraged to use the same special way that was previously heard. This newly added recall component served as practice to further help participants understand what it meant to remember the descriptions and recall them in the special way after a delay before the test trials began.

After this training portion, participants began the three experimental blocks (a general layout example is provided in Figure 1). As mentioned in the Design section, the three blocks could appear in one of six different orders, one of which being visual-spatial, then auditory-verbal, then no load, which will be described here. The visual-spatial block proceeded in the same manner as Study 1, with a few key changes. First, participants began with 4 practice trials of the load only. For the first practice trial, the 3 x 3 grid would appear with 2 tokens in two separate boxes for a total of 3000 ms. Then, a fixation point would appear for 2000 ms. Finally, a blank 3 x 3 grid would appear. Participants could either click or tap the boxes to cause a token to appear, whether their response was correct or incorrect (for participants on a non-touchscreen computer, it was recommended that they point to the boxes and allow their parents to click the boxes for them). After selecting all the boxes believed to contain a token during the probe, they clicked a button to indicate completion. This sequence occurred for a total of 3 more times during this load-only practice portion.

Next, participants were told they would continue to play the “token game,” but first, they would listen to descriptions of new images. Each image was accompanied by an audio which described the transitive action with one passive voice sentence. Participants were asked to repeat each image description and press an [I said it!] button before proceeding to the next. This occurred for a total of 8 trials. Then, the memory trials began. Participants were told they would play the token game again, but this time, an image would appear after they saw the tokens. After describing the image, they would be asked to recall the location of the tokens seen before the image. To help participants understand this sequence, I included one practice trial on this dual task, which included an image from the very beginning of experiment (i.e., one of the images they saw in the first training portion). Each trial proceeded as such: (1) a grid with 2 or 3 tokens displayed for 3000 ms, (2) an image whose description was previously heard and repeated appears and the participant is asked to recall the sentence, (3) the empty grid appears, in which the participant can indicate recalled locations of tokens. This trial sequence occurred for a total of eight times after the first practice trial.

The auditory-verbal load block also proceeded similarly to Study 1, except participants were asked to repeat each digit load immediately after hearing the audio from the experiment. In this version, participants began with 4 practice trials of the verbal load only. They were instructed that when they saw a green ear, they were to listen to the numbers spoken and repeat them after it finished. Then, when they saw a green ear with a yellow smiley face, they were to try to say the numbers again (requiring memory). The practice trials then followed such that participants would first see a green ear for 2000 to 3000 ms depending on the set size of the digit load and they heard the digits listed by a female voice. Then, a fixation point appeared for 2000 ms. Finally, the green ear and yellow smiley face combination appeared and participants were instructed to try to say those same numbers again. This sequence occurred for a total of 4 times.

Next, participants were told they would continue to play the “number game”, but first, they would listen to new descriptions of images. As with the visual-spatial block, they listened to and repeated 8 image descriptions. After repeating each description, the trials began with one practice trial which utilized a picture from the first training portion of the study. Each trial of this block proceeded as such: (1) A green ear would appear and the experiment would automatically play a recording of a female voice stating 2 or 3 numbers at a rate of about 1 s/digit. The participant immediately repeats the digits and then presses an [I said it!] button, (2) an image whose description was previously heard and repeated appears and the participant is asked to recall the sentence, (3) the green ear and yellow smiley face appear and the participant is to try to recall the same digits that were said before the image description. This sequence occurred for a total of eight times after the first practice trial of the block.

Finally, the no load block simply asked participants to listen to eight new image descriptions and repeat each. Then, they saw the same images and were asked to describe what occurred in each. There was no practice portion for this block, as it was an identical task to the training portion at the beginning of the experiment session. After the completion of these three blocks, the experiment was complete. Most participants completed the study in around 30 minutes, with some taking about 45 minutes.

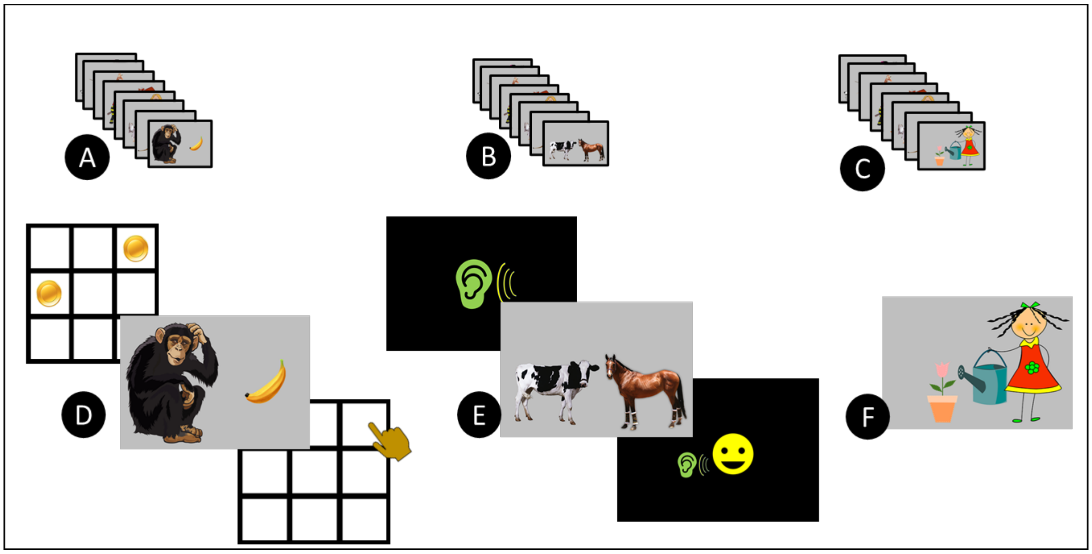


Figure 1. Experiment Layout. At the beginning of each block, participants would hear 8 passive-voice descriptions and repeat each one (A-C). In the visual-spatial block (D), participants would see tokens, see and describe an image, then tap or click boxes to indicate memory for token locations. In the auditory-verbal block (E), participants would listen to a short list of digits and repeat them, then describe one of the previously seen images, then say the numbers again. In the block with no load (F), participants simply described each previously seen image.

Results

Statistical Analyses

For all results, I use Bayesian analyses for inferential purposes and frequentist measures as complementary descriptive purposes. For overall effects, I report BF_{10} values as well as accompanying F values and effect sizes (η_p^2 or Cohen's d.) I use conventional cut-offs for BF_{10} values which specify values of 3.00 as denoting moderate evidence in favor of alternative hypotheses—the strength of this evidence is considered to increase as the factor value increases. For analyses that include more than one variable, BF_{incl} values are reported to denote whether terms were useful for keeping in the overall model. For correlations, I report Pearson's r values with accompanying BF_{10} values. All analyses were carried out using R (R Core Team, 2014) and JASP statistical software (JASP Team, 2020).

Passive vs. Active Use Overall

I began by assessing if participants used the passive or active more in their overall responses. Given a mean difference of 0.71 and standard deviation of 0.26, the data exceeded the minimum criteria for detecting an effect of a difference between passive and active voice use ($t(37) = 17.059$, $d = 2.76$, $BF_{10} = 1.18 \times 10^{16}$). There was strong evidence that participants used the passive voice ($M = 0.81$, $SE = 0.02$) more often than the active voice ($M = 0.10$, $SE = 0.02$). In fact, these rates were quite different from Study 1, with participants using the passive voice even more in Study 2 than Study 1 ($M = 0.65$; $BF_{10} = 13.63$) and the active voice less often than Study 1 ($M = 0.26$; $BF_{10} = 28.46$). This gave good reason to believe that the modifications in the method encouraged participants to try to use the passive voice more often.

There were other responses that did not fall into either the passive or active category. These could be responses in which the participant could not remember the description or said a response that was not clearly a passive or active voice sentence (e.g., simply labeling the patient and the agent). Overall, these types of responses comprised of 8.7% of total responses and did not vary by block order or load type, so I only refer to passive and active responses for the remainder of these analyses. There was no effect of block order on passive or active voice use. There was also evidence in favor of the null for an effect of load size on either passive or active voice use, so the remaining analyses are collapsed across load set size.

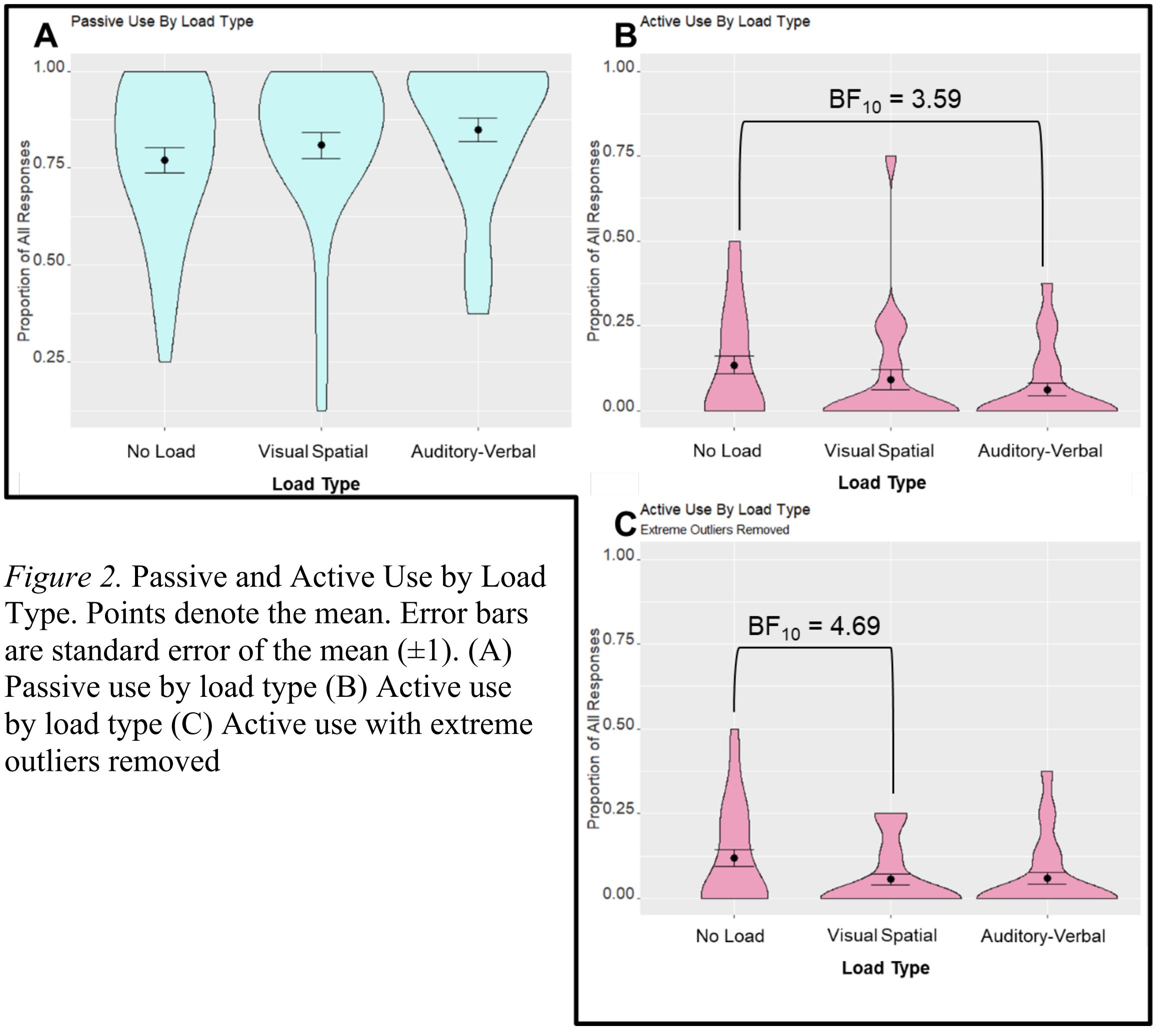


Figure 2. Passive and Active Use by Load Type. Points denote the mean. Error bars are standard error of the mean (±1). (A) Passive use by load type (B) Active use by load type (C) Active use with extreme outliers removed

Passive vs. Active Use by Load

Next, I assessed how passive and active voice use varied by the type of load participants were asked to hold using one-way repeated measures ANOVAs (depicted in Figure 2). As noted, with the sample size of 38, and the range of standard deviations when comparing these conditions (0.22 – 0.25), mean differences of about 0.10-0.11 are detectable and sufficient to achieve a Bayes factor of 3.2. For passive voice use, the mean differences between conditions did not reach this minimum mean difference and ranged from 0.04 to 0.08, making evidence for an effect of load type inconclusive (F(2, 74) =

2.15, η_p^2 = 0.06, BF_{10} = 0.48). Participants used the passive voice at similar rates whether they were holding a visual-spatial load (*M* = 0.81, *SE* = 0.03*),* verbal load (*M* = 0.85, *SE* =0.03), or no load at all (*M* = 0.77, *SE* = 0.03).

Considering the mean differences between conditions when participants were using the active voice, standard deviations ranged from 0.16 to 0.18, roughly. Mean differences of about 0.07-0.08 became detectable with the current sample size. Because most of the mean differences between these conditions fell just under or on the borderline of these minimums (0.03 – 0.07), there was also inconclusive evidence for an overall effect of load type on active voice use rate (F(2,74) = 3.46, η_p^2 = 0.085, BF_{10} = 0.752). However, post-hoc comparisons of active voice use between the different load types revealed moderate evidence for one effect: participants were more likely to use the active voice when there was no load (*M* = 0.14, *SE* = 0.03) as compared to when there was a verbal load (*M* = 0.06, *SE* = 0.02; BF_{10} = 3.59). There was no strong evidence for a difference between these two conditions and the visual-spatial load (*M* = 0.09, *SE* = 0.03).

For both passive and active use, I performed an outlier analysis using the **identify_outliers** function in the rstatix package in R (Kassambara, 2020). This analysis highlights all outliers, but denotes outliers that are 3 times the interquartile range of the 1st and 3rd quartile of the data as being extreme. For passive voice use by load type, there were no extreme outliers, so these

points were left in the data. However, for active voice use, two points in the data were considered extreme (both indicated extremely high usage of the active voice while under visual-spatial load). I removed these outliers to assess any changes in the outcome and found that there was now moderate evidence for an effect of load type on active voice use (BF_{10} = 3.70). Participants still spoke in the active voice the most when not under load, but specifically compared to when there was a visual-spatial load (BF_{10} = 4.69). With these outliers removed, the mean of active voice use under no load slightly decreased (from *M* = 0.14 to 0.12) and evidence for a difference between no load and auditory-verbal load active voice use became inconclusive (BF_{10} = 1.68).

Passive vs. Active Use by Load and Reversibility

In addition to assessing load type, I factored in the reversibility of the sentence by conducting two-way repeated measures ANOVAs with both factors included. First, I assessed how passive voice use was affected when both load type and reversibility were considered in the model. As before, there was inconclusive evidence for an effect of load type on the proportion of passive voice use. I also found evidence in favor of the null for an effect of reversibility ($F(1,37) = 1.25$, $\eta_p^2 = 0.03$, $BF_{10} = 0.25$). Participants used the passive voice at similar rates whether the sentence was reversible (*M* = 0.83, *SE* = 0.02) or irreversible (*M* = 0.79, *SE* = 0.03). There was also evidence in favor of the null for including an interaction term in the model (BF_{incl} = 0.048).

A similar story was revealed for active voice use when considering both load type and reversibility. There was inconclusive evidence for an effect of load type, as seen before. There was also inconclusive evidence for an effect of reversibility on active voice use ($F(1,37) = 3.92$, $\eta_p^2 = 0.10$, $BF_{10} = 0.50$), with participants using the active voice at similar rates whether the sentence was reversible (*M* = 0.08 , *SE* = 0.02), or irreversible (*M* = 0.12, *SE* = 0.03). This pattern of results remained true even with outliers removed.

There was also evidence in favor of the null for including an interaction term in the model ($BF_{incl} = 0.16$).

Errors in sentence recall

There were some instances when participants recalled the sentences incorrectly. These errors could come in different forms: direct reversals (switching the patient and agent of the sentences), incorrect labeling (e.g., calling a giraffe a zebra), or a combination of these different types of errors. For the analysis here, I combine all types of errors. In the supplement, I include a more detailed breakdown of all the observed categories of errors and their rates of occurrence.

Overall, participants produced errors at a relatively low rate ($M = 0.09$, $SE = 0.02$). The rate of error decreased slightly from Study 1 ($M = 0.14$, $SE = 0.01$), but there was inconclusive evidence of a difference in error rate between the two studies ($BF_{10} = 0.40$). Most errors were made up from attempts to use the passive voice ($M = 0.08$, $SE = 0.02$) and the remaining were active voice errors ($M = 0.01$, $SE = 0.00$). There was inconclusive evidence for a difference between rate of errors within only passive voice use ($M = 0.10$) and only active voice use ($M = 0.15$; $BF_{10} = 0.26$). Error rates also did not vary by load type or reversibility, except for one result. Overall, participants were more likely to produce *correct* active voice sentences when the sentence was not reversible ($M = 0.11$) as compared to when the sentence was reversible ($M = 0.06$; $t(1,37) = 2.176$, $BF_{10} = 4.98$), a result that mirrored Study 1. All other effects were null or inconclusive (specific breakdown provided in Appendix B).

Load performance and its effects

The key manipulations in these studies are the working memory loads that participants were required to hold in two of the three blocks. In order to better understand how participants were using their working memory, I analyzed performance accuracy on these load tasks and how it affected performance on the language task. For each load type, participants completed 4 trials used as practice without the additional language task. Then, they completed 8 trials with the language task. I first analyzed overall performance for these different trials (the accuracy data for two participants could not be included because of data recording errors). For visual-spatial loads, participants performed at higher rates of accuracy during the practice trials (M = 0.77, SE = 0.04) as compared to the dual-task trials (M = 0.49, SE = 0.03, t(35) = 6.36, d = 1.06, BF_{10} = 6.00 x 10^4). For auditory-verbal loads, participants also performed better during the practice, single-task trials (M = 0.93, SE = 0.03) than the dual-task trials (M = 0.42, SE = 0.03, t(34) = 12.91, d = 2.18, BF_{10} = 5.85 x 10^{11}; 3 participants' data were not included for the practice auditory-verbal trials due to data recording errors). There was indeterminate evidence for a correlation between accuracy on the single-task trials and dual-task trials for both the visual-spatial (r = 0.30, BF_{10} = 0.94) and auditory-verbal (r = 0.36, BF_{10} = 1.91) load conditions.

Although practice trial performance on the auditory-verbal load was much more accurate than practice trial performance on the visual-spatial load (Mean difference: 0.16, BF_{10} = 209.06), performance during the dual task trials were more comparable between the two load types and even correlated (r = 0.480, BF_{10} = 14.13). Performance on the load task did not predict whether participants used the passive voice (r = -0.02, BF_{10} =

0.20) or active voice (r = 0.11, BF_{10} = 0.25). It also did not predict whether participants would produce an error (r = 0.22, BF_{10} = 0.47) or not (r = -0.14, BF_{10} = 0.28).

Discussion

In this study, I sought to find if emphasizing to participants that they must use the target syntax, the passive voice, and requiring more sentence repetitions would change our pattern of results from Study 1. Overall, I found that results were similar to that of Study 1, although some results became inconclusive or in favor of the null. Next, I will revisit each of these key findings from Study 2:

(1) Participants used the passive voice much more often than the active voice, and also at higher rates than was found in Study 1.

(2) While there was inconclusive evidence for overall effects of load type on either passive or active voice use, I found post-hoc evidence that participants utilized the active voice more often when there was no load, as compared to when there was a verbal load. This becomes an overall effect when extreme outliers are removed, and active voice use is greater in the no load condition as compared to the visual-spatial condition.

(3) Participants produced errors at a similar rate to Study 1, but these errors did not vary by load.

More passive voice use and less active voice use

The finding that participants' rate of passive voice use increased and active voice use decreased from Study 1 rates was encouraging. Two key changes likely led to these heightened effects: more training and forced repetition. The modified method of including more information about the passive voice might have served as a better background for what the “special way” sounds like beyond simple repetitions. In Study 2, participants were not only able to hear and practice the passive voice, but also listen to examples of what is *not* the passive voice. I was able to use responses from Study 1 to note two of the most common alternative sentence types (i.e., active voice and direct reversal passives). I also added a delayed recall portion to the practice, further helping children to understand that they would be asked to hold the passive voice sentences in mind before recalling them later on. In the first study, the practice simply involved repeating passive voice sentences, which may have helped to denote what the passive voice sounded like, but might not have prepared participants for the possibility of having to remember the sentence. Thus, the new training portion at the beginning of Study 2 may have contributed to these higher rates of passive voice use and drop in active voice use.

Another possible factor is the new requirement for participants to repeat each probe sentence before continuing. In Study 1, we did not ask participants to repeat the probe sentences and they were informed that they could simply listen as the experimenter described each image, one at a time. Now in Study 2, participants were required to say the sentence immediately after each probe before continuing to the next. This repetition may have functioned much like rehearsal for participants. Repeating each sentence not only ensured that participants heard it fully and correctly at the probe, but helped them to practice saying the sentence in the correct way. Thus, repeating the sentence again, even after a delay, was likely easier and the sentence itself may have felt more familiar.

More active voice use when there was no load

One of the key results from Study 1 was that participants were more likely to use the active voice when there was no load. Though evidence for effects were not as robust as Study 1, I found a similar pattern in Study 2 results in post-hoc comparison. In this study, participants were more likely to use the active voice when there was no load as compared to when there was a verbal load, only. Moreover, when two extreme outliers were removed there was a main effect of load, and the effect of a difference between the no-load versus visual load conditions became reliable. These results suggest that, though participants were able to significantly drop their use of the active voice, they still preferred to do so more often when there was no load. Replicating this specific effect from Study 1 builds more evidence for the conclusion that paraphrasing sentences to a more familiar (or just different) syntax while maintaining the meaning requires some level of working memory. This process likely takes a higher level of comprehension of the passive voice sentence and some central executive processes to transform the sentence correctly. For children, avoiding errors in speech, especially when dealing with a relatively unfamiliar syntactic form, requires free working memory.

It could also be that the sentence repetitions reduced the need for attention-based working memory when using the passive voice. Working memory is especially useful for handling information before it is rapidly learned and added to the long-term memory system (Cowan, 2019). It could be that the probe repetition requirement allowed for rapid learning—adding the sentence elements to activated long-term memory. If this is the case, the need for attention-based working memory could become less important when the sentence is being recalled. However, attention may be re-recruited when the speaker wants to use a syntax other than the one that was already produced during the repetition phase. This would explain the greater likelihood of transforming the sentence to the active voice while under load. If attention is occupied with other information, it is best to use the sentence available in activated long-term memory which does not require as much attention, hence the passive voice use on the majority of trials.

Error rates

After modifying the methods of Study 1 for this second study, I hypothesized that emphasizing the need to attempt to use the passive voice would elicit even more errors for children. The goal was to emphasize syntax over semantics and increase the number of semantic errors, overall. It is possible that, although participants used the passive voice more often, their level of comprehension was not any higher than it was for Study 1. This is suggested by the similar rate of errors between the two studies. Additionally, although the majority of errors in Study 2 were made up of attempts to use the passive voice, each of the rates of errors with the passive and active voice were similar. It could be that the new requirement to repeat each sentence served as a protective factor against higher rates of passive voice errors than active voice errors. In Study 1, which contained no repetition requirement, 17.4% of passive responses were errors while only 8.2% of active responses were errors ($t(30) = 2.44$, $BF_{10} = 4.76$).

What appeared to still matter to some extent was the nature of the sentence itself. Another effect which replicated from Study 1 was the higher likelihood of using the active voice correctly when the sentence was *not* reversible (e.g., the balloon was popped by the giraffe). This could be because there is only one logical way to interpret the action in the image or sentence, which makes it easier to transform the sentence to another syntax without sacrificing semantics. This result provides more evidence for the importance the level of ease in comprehending a sentence for subsequent reproduction.

In sum, participants produced a similar rate of overall errors and a similar rate of passive voice errors from Study 1 to Study 2. However, in Study 2, there was no difference in rates of passive voice errors as compared to active voice errors, as there was in Study 1. It could be that sentence repetitions serve as a buffer against higher rates of error production, even when the passive voice is being used significantly more often. It can also reduce the need for attention-based working memory which allows for rote learning and a higher likelihood exactly as presented previously.

Conclusions and next steps

In Study 1, we found that participants were using the passive voice more often, but still relied on the active voice about 26% of the time, and especially when there was no load. We also found that participants were more likely to produce errors if they used the passive voice, and these passive voice errors seemed to especially appear when there was a working memory load. We hypothesized that children may have been more concerned with portraying the correct *meaning* of the sentence, rather than saying it in the target syntax, which is easiest to do when there is no additional information to hold in working memory. Thus, for Study 2, I attempted to emphasize the importance of using the passive voice by adding more training and having participants repeat probe sentences as they heard them. I found that these changes effectively increased children's use of the passive voice and decreased their use of the active voice. However, they also produced a similar overall rate of errors and working memory loads did not affect error production.

I found the results of Study 2 to be intriguing enough to move forward with the next challenge for 4- to 5-year-old speakers: producing their own original passive voice sentences. Given that method changes increased instances of the passive voice, yet error rates did not increase, one way to be able to tell if participants actually *understand* the "special way" is to require them to create their own constructions. If able to do so correctly, this would be an ultimate test of participants' ability to understand and use a relatively novel construction and to assess how working memory contributes to this ability. Though the method of recalling sentences has given us a starting point for understanding how children deal with unfamiliar language, asking them to use the unfamiliar language equipped with their own vocabulary and interpretation of events is more like language acquisition tasks in the real world. In the next chapter, I discuss the study that attempted to do just this.

Chapter 5. Study 3: Creating New Passive Voice Constructions

For the young speaker, some language production may come in the form of trying to reproduce words and phrases heard by other speakers around them. However, much of language requires us to produce new sentences about objects, people, or situations we have not encountered before. Very young speakers must take on the task of trying to use words and syntactic structures that are new or still developing to accurately discuss ideas and events. In this final study, I attempt to recreate this latter situation by having children create their own passive voice sentences about transitive-action animations. I also assess how working memory contributes to the ability to create these passive voice sentences by requiring participants to do so under working memory load, as in the previous studies. Given the number of factors that could affect children's ability to do this challenging task, auxiliary measures were added to test how children's working memory span and language abilities serve as predictors.

Images versus animations

Our first goal was to devise a method in which participants could view an image and be able to create their own passive voice sentence about it after brief training. This method would be more akin to that of Huttenlocher et al. (2004) in which participants listened to an experimenter describe 10 images in the passive voice and were then asked to describe 10 images on their own (with no direct emphasis on the syntax). Children who heard the passive voice were 14% more likely to use the passive voice in their own sentence description. In that study, images were drawings displaying a specific action (e.g., a rabbit being chased by a dog) so that children could interpret the transitive verb and the displayed nouns with relative ease. However, in our stimuli in Studies 1 and 2, the images were purposely crafted so as not to display a specific verb (requiring participants to rely on their memory of the sentence rather than interpretation of the image). Although children were able to recall the main content of each image description at very high rates, attempting to produce a new sentence about these images would potentially present new problems. For one, because there was no distinct verb action being displayed in the images, it was possible that participants would not be able to easily create their own full, transitive-action sentences, but instead would resort to simply naming the nouns displayed (e.g., *It's a flower and a girl*) or describing a non-transitive action (e.g., *The horse and the cow are walking together*). This also left open the possibility of differences in sentence production occurring because of variations in vocabulary size, creativity, talkativeness, etc.

Thus, in order to clearly portray a transitive verb taking place between a patient and an agent, I transformed each image into a video animation. Now, the agent and patient of each sentence could move and interact in order to display the intended verb. Even with a verb displayed in an animation, it was expected that there would still be a degree of variation in interpretation and words chosen (e.g., one child might prefer to say *jumped on*, while another is more familiar with the verb phrase *hopped on*). Our only requirement was that children try to use the passive voice when describing the animation, whether the verb matched the one that was originally intended to be portrayed or not. In order to test if children would understand that there was a transitive action taking place, I piloted each animation with a small group of children before testing and tuned animations that were less clear. I attempted to maintain the original sentences from Studies 1 and 2, although some images were replaced with new sentences and animations that were deemed easier for viewers to discern the action.

Hypotheses

Now that participants were required to construct their own, original passive voice sentences about animations that had not been previously witnessed, I expected a few key changes in results. First, I expected that participants would use the active voice more often than the passive voice, given the difficulty with using the passive voice. It could have also been possible that participants would use the passive voice more often because brief training is adequate enough to encourage use for the majority of descriptions.

Second, I expected that participants would rely on the active voice the most when trying to hold additional working memory loads and the most passive voice use would occur when there was no load. This result would imply that forming sentences in an unfamiliar syntax is what requires the most working memory and would also help frame how participants were simple repeating the passive voice sentences in Studies 1 and 2, a task that is much different than constructing new sentences. However, it was alternatively possible that participants' responses would not be affected by working memory loads. One possible explanation for this sort of result would be that constructing new sentences, whether in the active or passive voice, is a demanding enough task that working memory loads are ignored no matter the syntax used.

One way to test the impact of working memory loads on the language task outside of observing choice of syntax is to compare performance on practice trials in which participants remember only the load (single-task) and the test trials in which participants remember the load while constructing the passive sentence (dual-task). In

Study 2, performance on the load task decreased in accuracy from the single- to dual-task trials. If, in Study 3, constructing sentences requires an even greater amount of working memory, it is expected that the decrease in accuracy from single- to dual-task trials will be greater. If, however, these processes do not depend on the same attention resource, performance on the tasks may appear similar – whether the participant is trying to construct a passive sentence or not. This latter possibility is in line with the views of Waters & Caplan (1999), who posit that processing syntax requires a separate and special working memory resource than other types of information (to be revisited in the Discussion section).

Third, and last, I expected that participants would produce even more errors in the passive voice than in Studies 1 and 2 because of the lack of a model sentence to repeat. Alternatively, participants could produce less errors in the passive voice because they choose to only use the passive voice if they are confident their construction matches their intended meaning. In sum, the results from Study 3 will help frame the depth of children's understanding of the passive voice and give more clues as to why they choose to use one syntax over the other.

Method

Participants

For Study 3, I recruited participants through Facebook advertising, as well as digital ads sent to parents with children within the local school district, and to the university at large. Thirty-six participants (23 male, 13 female) ranging from 4:0 to 5:9 (MA: 4:10) completed the study with minimal distractions over the course of two sessions (one participant did not return for the second session). Now, with a sample size of 36, and the selected standard deviations of this study (0.22 to 0.55), mean differences of 0.10 to 0.30 should be detectable and sufficient for a conclusive Bayes factor, depending on the measure (as calculated by the BayesFactor package in R). I re-address these criteria in selected analyses in order to assess the power of the current sample size. Of the completed studies, 26 participants were Non-Hispanic White, 4 were Asian, 2 were Hispanic White, 1 was Black or African American, 1 was American Indian or Alaskan Native, and 2 did not provide ethnic or racial information. All children had normal or corrected vision. One child was reported as not having normal hearing at the time of testing. Three children were reported as having a first language other than English, and two of these parents indicated less than fluent English speech. For these children, I proceeded with the experiment and found no significant differences between their native-English speaking peers on study performance and were thus maintained in the sample.

Design

As in Studies 1 and 2, I continued to use the three-block design in which participants were required to a hold a working memory load for two of three blocks. However, the beginning portion of the experiment which introduced the passive voice changed because the task would no longer require recall, but construction of original sentences. The order of the three experimental blocks after this training block were still counterbalanced so that participants could receive one of six different overall orders. The videos were grouped into three groups as before, with half of each group containing reversible stimuli and the other half containing irreversible stimuli. The order of the images was also counterbalanced in three different orders, as before, so that there were 18 possible experiment orders. Within each block, the order of the loads and animations were completely randomized by the experiment software. As in Study 2, some animations were yoked to certain load sizes, so that it was not possible for one participant to see a video under a load size of 2 and another under a load size of three. Any given video would only appear with one load size (either 2 or 3) or no load at all.

Stimuli

A total of 30 animations were created, 6 of which were used in the training and practice portion of the experiment, and the remaining 24 were used for the test trials (sentence stimuli and a link to all animations are available in Appendix C). In order to create each animation, I used the stimuli from Studies 1 and 2, but altered them using Adobe Photoshop and Microsoft PowerPoint. The images were then animated using PowerPoint, where the final videos were created. Each animation lasted from roughly 3 – 9 seconds depending on the action taking place, but the transitive action was discernible within roughly 3 seconds of the video beginning. For example, an animation of a dog chasing a cat took longer because the dog chased the cat around the screen display for several seconds. However, the chasing began as soon as the video started. Videos were counterbalanced so that half of agents were depicted on the left, and the other half on the right, with a couple exceptions in which the agent was holding the patient. Each video was played on a loop so that participants could view the action repeatedly until they devised a descriptor sentence. Each video was displayed on a light gray background superimposed on a larger white background so that the target action was especially visible and highlighted on the gray background. Videos contained no audio.

In order to check if each animation portrayed the intended meaning, I conducted a pilot study with a small sample (n = 7) of 4- and 5-year-old children. The pilot study consisted of watching each animation and simply stating what they interpreted as occurring, with no specific instructions on syntax. If the majority of children were able to discern the same meaning from the videos, the videos were kept. If responses were unrelated or responses did not match the intended message, the video was modified. Only 3 videos matched the latter criteria and were modified to be included in the final stimuli. The final portion of the pilot study included a "special challenge" in which I tested whether children were able to use the special way to describe three videos after brief training. Because multiple children were able to produce the passive, I proceeded with the training method. The pilot study lasted about 10 minutes total for all participants.

The load stimuli remained identical to Study 2, using 3 x 3 grids with gold tokens for the visual-spatial loads and 2 or 3 digits used for the auditory-verbal loads. I also added two auxiliary measures for Study 3: the Quick Interactive Language Screener (QUILS; Golinkoff et al., 2016) and a forward auditory digit span test. The QUILS is an online screener designed for children 3- to 5-years-old. The test consists of 48 items that span three content areas: vocabulary, syntax, and process. The vocabulary area tests children's knowledge of nouns, verbs, prepositions, and conjunctions. The syntax area tests children's mastery of Wh-questions, the past tense, prepositional phrases, and embedded clauses. The process area tests how well children can learn new words and use syntactic context to learn new vocabulary and structures. This section tests verb learning, converting the active to the passive, noun learning, and adjective learning. Standard scores and percentile rank for each sub-part, as well as the overall test are available to the experimenter immediately after testing.

The auditory digit span consisted of groups of 3 lists of digits of increasing set size spoken aloud by a female voice that was recorded and presented to participants through online experiment software (Psytoolkit) at a rate of 1 digit/s. The experimenter provided verbal instructions and entered responses for each participant into the experiment interface.

Procedure

This study was completed in two sessions. By default, the first session was reserved for completing the main experiment and the second session was for completing the digit span and QUILS tests. For a small subset of the participants (n = 7), the experiment malfunctioned so that it could not be completed in the first session. For these participants, the auxiliary tests were completed during the first session and the experiment in the second. All participants were asked to return within two weeks of the first session. For the experiment session, as in Study 1, each participant's parent or guardian was provided with a link to the online experiment on the Psytoolkit platform. After obtaining parental consent and child assent, obtaining demographic information, and working through technical logistics, the experiment began.

For Study 3, I attempted to increase the motivation to try to use the special way by presenting participants with a background story in which they were traveling to a different planet (experiment graphics displayed an alien character in outer space). The scenario was that on the planet, the inhabitants spoke in a "special way" (i.e., still the passive voice) and the goal was to try to speak in the same manner. The experiment then began with the same extended example as Study 2, now displaying an animated version of the sun and melting ice cream. Participants were provided with a repetitive example of a correct passive voice sentence (*The ice cream was melted by the sun*), and also examples of other sentence forms that were not the special way. Participants were asked to repeat the correct passive voice sentence after the experimenter. After the extended example, two additional animations were displayed, one at a time, and the participant repeated the passive voice description after the experimenter. Then, participants were told they would attempt to create their own passive voice descriptions of 3 new videos.

For each practice trial, participants were shown an animation and asked to describe it using the special way. If the participant uttered an active voice sentence or any other variation of a response outside of a passive voice sentence, the experimenter first began by trying to elicit the passive through a passive-inducing question (e.g., *What happened to the ball?*). If the participant still did not use the passive voice, the experimenter restated the sentence in the passive voice and asked the participant to repeat the sentence again in the "correct" way. Once the child uttered a passive voice description about the animation, they could proceed to the next. After describing the three practice videos, the training and practice block was complete. The rest of the experiment was divided into three blocks (presented as rounds to the participant), the order of which was randomized as described in the Design section.

A major difference in the method of Study 3 as compared to Studies 1 and 2 was that there was no recall requirement for the sentence stimuli. Animations seen during the test trials were all brand new to the participant and the sentence they constructed was required to be completely original. For the visual-spatial block, participants began with the same four practice trials of the load task on its own as in Study 2. Then, they were notified that they would continue to play the “token game,” but also be asked to talk about new videos in the special way. The first trial was for practice of the dual task and followed as such: (1) the probe grid was displayed with 2 tokens dispersed for a total of 3000 ms, (2) a video animation appeared (for the practice trial, this was a video repeated from the training portion) and the participant was asked to describe it in the special way, and (3) an empty grid appeared and participants clicked or tapped on each box where they remembered seeing tokens. This same sequence repeated for 8 more test trials, with set sizes varying between 2 and 3 tokens, all with new animations for the sentence task.

The auditory-verbal block also began with four practice trials, identical to those of Study 2. After the practice trials, participants were told they would continue to play the “number game,” but also describe a new set of videos in the special way. There was one practice trial of the dual task which followed as such: (1) A green ear appeared and audio of a female voice stating 2 digits at approximately 1 digit/s and the participant was asked to immediately restate the short list of digits, (2) an animation appeared (for the practice trial, this was a video repeated from the training portion) and the participant was asked to describe it in the special way, (3) A green ear and yellow smiley face appeared and the participant was asked to restate the numbers from the probe. This same sequence was repeated for 8 more test trials, with set sizes varying between 2 or 3 digits, with new animations for each sentence portion of the task.

In the final block, the participant was simply asked to describe 8 animations in the special way with no additional memory task. After the completion of all three experimental blocks, the session was complete. Each session was audio recorded and the majority were also live-coded as the participant provided response. The experiment session took an average of 30 minutes total for participants to complete, including form completion and setup time.

Auxiliary measures

In a separate session, participants completed the auditory digit span and QUILS. The auditory digit span required participants to listen to groups of 3 lists of digits increasing in set size (beginning at set size 1). The participant was asked to wait until the completion of each list to repeat the digits in the same order as heard. The experimenter then entered each response. The experiment concluded when the participant responded incorrectly to all 3 lists within a block of lists. An incorrect response was any other response but the same exact list of digits.

The QUILS is an online assessment that participants complete by listening to questions about various pictures and videos and clicking pictures that match their desired answer. To complete the QUILS portion, the experimenter began the online program and shared the screen with the participant and provided remote control. The experimenter remained available for questions or general guidance, but the entire assessment is largely independent on the part of the participant. This second session of auxiliary measures typically lasted an average of 25 to 30 minutes.

Results

Passive versus active voice use

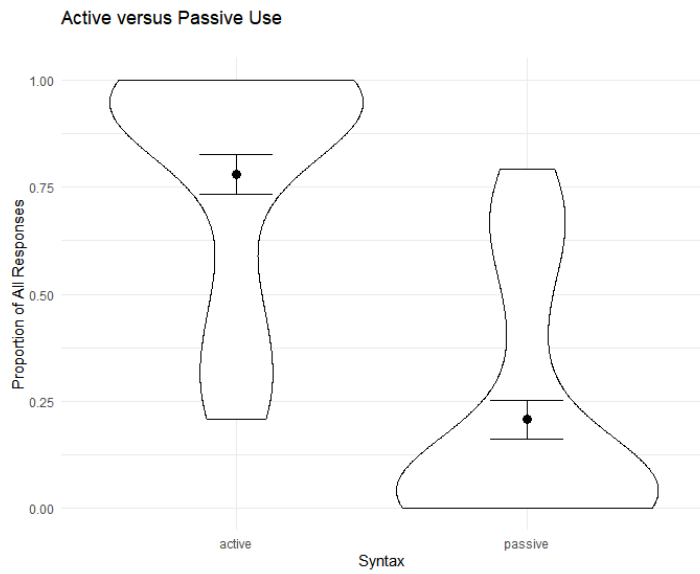


Figure 3. Passive versus Active Use Overall. Points denote mean averages. Error bars are standard error of the mean (±1).

My first analysis asked whether participants were able to use the passive voice at all in their original constructions about transitive action animations. Given a mean difference of 0.57 and standard deviation of 0.55, the data exceeded the minimum criteria for detecting an effect and provided extreme evidence of a difference between passive and active voice use ($t(35) = -6.220$, $BF_{10} = 4.09 \times 10^4$, $d = -1.037$). Participants used the active voice ($M = 0.78$, $SE = 0.05$) more often than the passive voice ($M = 0.21$, $SE = 0.05$; Figure 3). Overall, participants were able to describe the vast majority videos and only 1.2% of responses fell outside of the active or passive category (e.g., “I don’t know.”). Neither passive nor active voice use varied by block order. Though active voice use was extremely high, the finding that 21% of participants’ responses were in the passive voice after brief training was encouraging. 69.4% of participants were able to use the passive voice in at least one response, whereas the remaining participants used the active voice in all responses. 31% of participants were able to use the passive voice in at least 20% of responses and 22% used the passive voice in the majority of responses (illustrated in Figure 4). An outlier analysis revealed no extreme outliers in use of the passive or active voice in any of the three conditions.

There was also evidence in favor of the null for an effect of load size on either passive or active voice use, so the remaining analyses are collapsed across load set size.

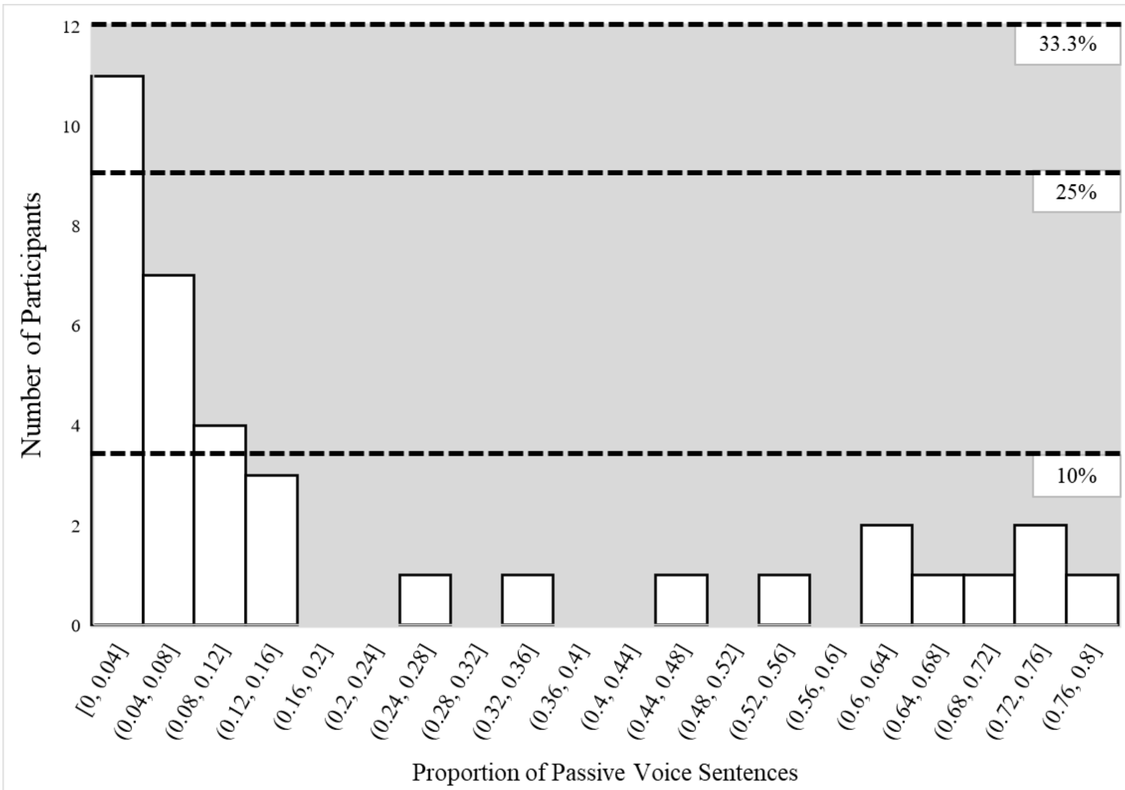


Figure 4. Histogram of participants' use of the passive. Binned into proportions of 0.04 each because 1 response in the passive voice is equivalent to 0.04 percent of responses.

I also conducted an item-level analysis to assess if certain animations were easier or more difficult to describe using the passive than others (breakdown included in Appendix D). Sentences that elicited low or high passive performance that fell at least one standard deviation above or below the mean were highlighted as being particularly difficult or easy to describe in the passive. Only five sentences fit these criteria: four sentences fell at least one standard deviation below the mean (indicating relative difficulty) and one sentence fell at least one standard deviation above the mean (indicating relative ease). I assessed the four difficult sentences for any similar elements and there were no obvious characteristics that made each of these sentences more difficult than the rest. The relatively easy sentence, however, contained the same action

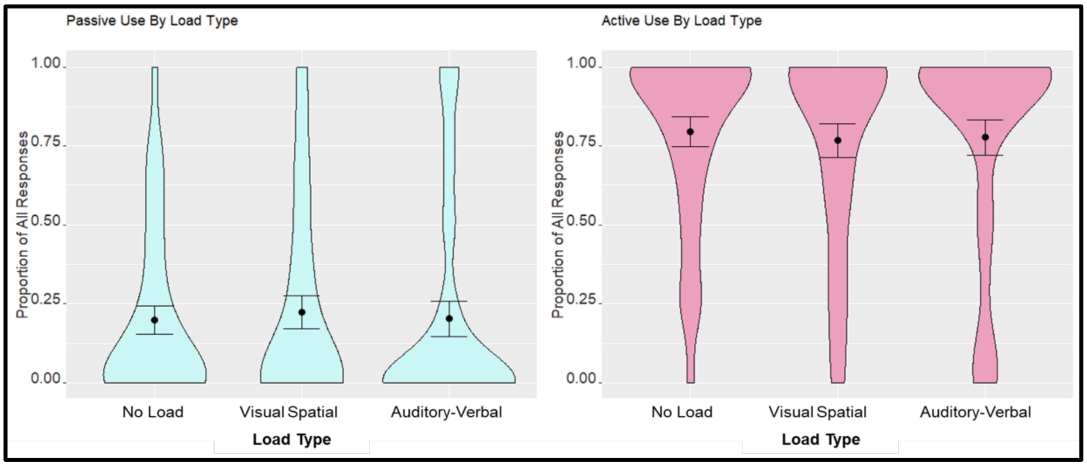


Figure 5. Passive and Active Use by Load Type. Points denote the mean. Error bars are standard error of the mean (±1).

as another animation used in the training block, which may have given participants an advantage towards using the passive voice. Given that the remaining 19 sentences elicited similar performance, all sentences were maintained in the remaining analyses.

Passive versus active voice use by load type

I performed one-way repeated measures ANOVAs to assess if load type affected passive and active voice use. There were not similar patterns of effects of working memory load to those found in Studies 1 and 2 (Figure 5). Now with standard deviations ranging from 0.21 to 0.30 for both passive and active voice use, mean differences did not reach the minimum criteria for detecting a difference syntax use between three different conditions. The likelihood of participants speaking in the passive voice was similar whether the sentence was being produced under a visual-spatial load (*M* = 0.22, *SE* =0.05), auditory-verbal load (*M* =0.20, *SE* =0.06), or no load (*M* = 0.20, *SE* = 0.05; $F(2,70) = 0.200$, $\eta_p^2 = 0.01$, $BF_{10} = 0.10$). The same was true for active voice use, in which there was also evidence in favor of the null for an effect of load type. Rate of active voice use was similar whether the load was visual-spatial (*M* = 0.77, *SE* = 0.05), auditory-verbal (*M* = 0.78, *SE* = 0.06), or if there was no load (*M* = 0.80, *SE* = 0.05; $F(2,70) = 0.242$, $\eta_p^2 = 0.01$, $BF_{10} = 0.10$). An outlier analysis revealed no extreme outliers in use of the passive or active voice in any of the three conditions.

Passive versus active use by load type and reversibility

Next, I performed two-way repeated measures ANOVAs with both load type and reversibility in the model. For Study 3, although the participant had not yet heard the description sentence, reversibility was still based on whether the patient and agent could be logically switched in the scenario. In this case, a boy cutting a sandwich follows logic, but not the other way around. However, a horse pushing a cow could also be reversed into a cow pushing a horse.

For passive voice use, there was inconclusive evidence for an effect of reversibility on production rate. Participants used the passive voice at similar rates whether the animation presented a reversible (*M* = 0.17, *SE* = 0.05) or irreversible (*M* = 0.23, *SE* = 0.05) situation ($F(1,35) = 6.78$, $\eta_p^2 = 0.16$, $BF_{10} = 1.00$). There was evidence in favor of the null for an interaction of load type and reversibility ($BF_{incl} = 0.02$).

A similar pattern of results emerged for active voice use, in which there was inconclusive evidence for an effect of reversibility. Participants used the active voice at similar rates whether the animation was reversible (*M* = 0.79, *SE* = 0.04) or irreversible

(*M* = 0.74, *SE* = 0.05; F(1, 35) = 4.28, η_p^2 = 0.11, BF_{10} = 0.59). There was also evidence in favor of the null for an interaction term (BF_{incl} = 0.01).

Errors in sentence production

The presence of errors in sentence production was unique in Study 3 because of the lack of a probe sentence. Now, because the participant was constructing their own passive voice sentence, the nouns and verbs used could vary widely, as long as it maintained logic. Each participant's response was judged by this standard and several common types of errors were revealed, although overall error rate was low (6.1%). I include a breakdown of the different error types in the Appendix B.

Overall, results showed that load type did not affect production of errors. One result that flipped from the results of Studies 1 and 2 was the syntax in which the most errors occurred. In Study 3, the majority of errors produced were in the active voice (*M* = 0.05, *SE* = 0.01), and the remaining minority were made up from passive voice responses (*M* = 0.01, *SE* = 0.00; BF_{10} = 66.56). However, as in Study 2, the proportion of errors within each syntax was comparable. Of all passive voice responses, 14.9% were errors and of all active voice responses, 8.9% were errors (evidence for a difference between the two rates was inconclusive BF_{10} = 0.593).

Load performance effects

Next, I assessed participants' performance accuracy on the working memory load task. For the visual-spatial loads, in the practice trials before the dual task, participants responded correctly around 60.5% of responses. This is a notable drop from practice trial performance in Study 2 (*M* = 0.77), and the cause is unclear given that this part of the task remained identical. As expected, performance on the visual-spatial load did decrease in accuracy during the dual task trials (*M* = 0.37, *SE* = 0.04), again a drop from Study 2 performance (*M* = 0.49), but there was inconclusive evidence for a difference (t(1, 34) = 1.58, BF_{10} = 1.16). For auditory-verbal load performance, accuracy on the practice trials was relatively high (*M* = 0.88, *SE* = 0.05), then dropped to a similar level of accuracy (*M* = 0.38, *SE* = 0.03) during the test trials as with the visual-spatial loads. Overall, even though accuracy rates initially appeared to decrease from Study 1, statistical evidence for a difference was inconclusive.

There was also inconclusive evidence for a correlation between performance on the load tasks on passive voice use (*r* = 0.26, BF_{10} = 0.65) and also between performance and active voice use (*r* = -0.24, BF_{10} = 0.55). When assessing the rare errors that occurred, although overall performance on the load tasks did not predict rate of errors in the passive voice (*r* = -0.14, BF_{10} = 0.29), it did predict whether participants would make an error in the active voice (*r* = -0.50, BF_{10} = 20.11). Participants were more likely to produce errors when they answered less accurately on the load tasks, the opposite of what might be expected. This effect was largely driven by performance on the auditory-verbal load, in particular. The better participants did at recalling the digits during the test trials, the less likely they were to make errors in the active voice, overall (*r* = -0.43, BF_{10} = 5.71).

Auxiliary measures

Digit span. Participants also completed an auditory digit span test as a measure of working memory capacity (9 participants did not complete the digit span due to time constraints or technical issues). Each participant was assigned a basal and ceiling score based on responses: basal scores equated to the last set size in which the participant repeated all 3 lists correctly and ceiling scores denote the set size in which the participant repeated all 3 lists incorrectly (thus concluding the test). On average, participants basal score was 3.03 digits. The average ceiling score was 4.85 digits. Typically, the basal score is used as the participants' digit span. This score did not predict participants' likelihood of speaking in the passive (r = 0.12, BF_{10} = 0.28) or active voice (r = -0.08, BF_{10} = 0.26). It also did not predict error production rates or performance on the load task.

QUILS. Participants also completed the QUILS as a measure of vocabulary, syntax, and processing (the act of acquiring new language). Overall, participants varied widely in percentile rank compared to their peers (27.90 to 99.90). The average percentile ranks for vocabulary were 74.66% (a range of 34.4 – 99.9), 75.2% for syntax (13.0 – 99.9), 69.8% for processing (26.7 – 99.9).

The QUILS provides an overall standardized score for each participant taking into account their age in years and months on the day of testing and combines their vocabulary, syntax, and processing scores. I assessed if this standard score predicted any performance in the experiment and found moderate evidence that participants with higher standard scores were more likely to be able to use the passive voice (r = 0.42, BF_{10} = 3.93). Furthermore, this correlation was especially driven by passive voice use when participants were under load, whether that was visual-spatial (r = 0.43, BF_{10} = 4.71) or auditory-verbal (r = 0.45, BF_{10} = 6.14). QUILS performance did not predict passive voice use when there was no load (r = 0.23, BF_{10} = 0.48). It was also not predictive of active voice use, in which the correlation was inconclusive (r = -0.37, BF_{10} = 2.00; Figure 6).

I also assessed the pattern of these effects for only the top QUILS performers by separating out participants whose percentile scores fell into the top third of all participants (n = 11). By isolating these data points, it is possible to ask if top QUILS performers were simply speaking in the passive voice more, overall, or if memory loads did affect when they were speaking in the passive. At first glance, the latter seemed true. These top performers used the passive voice around 44% of responses when under spatial visual load, 38% when under visual-spatial load, and only 21% of responses when there was no load. However, an analysis of effects yielded inconclusive evidence for an effect of load type on passive voice use amongst these participants (BF_{10} = 1.365; this small

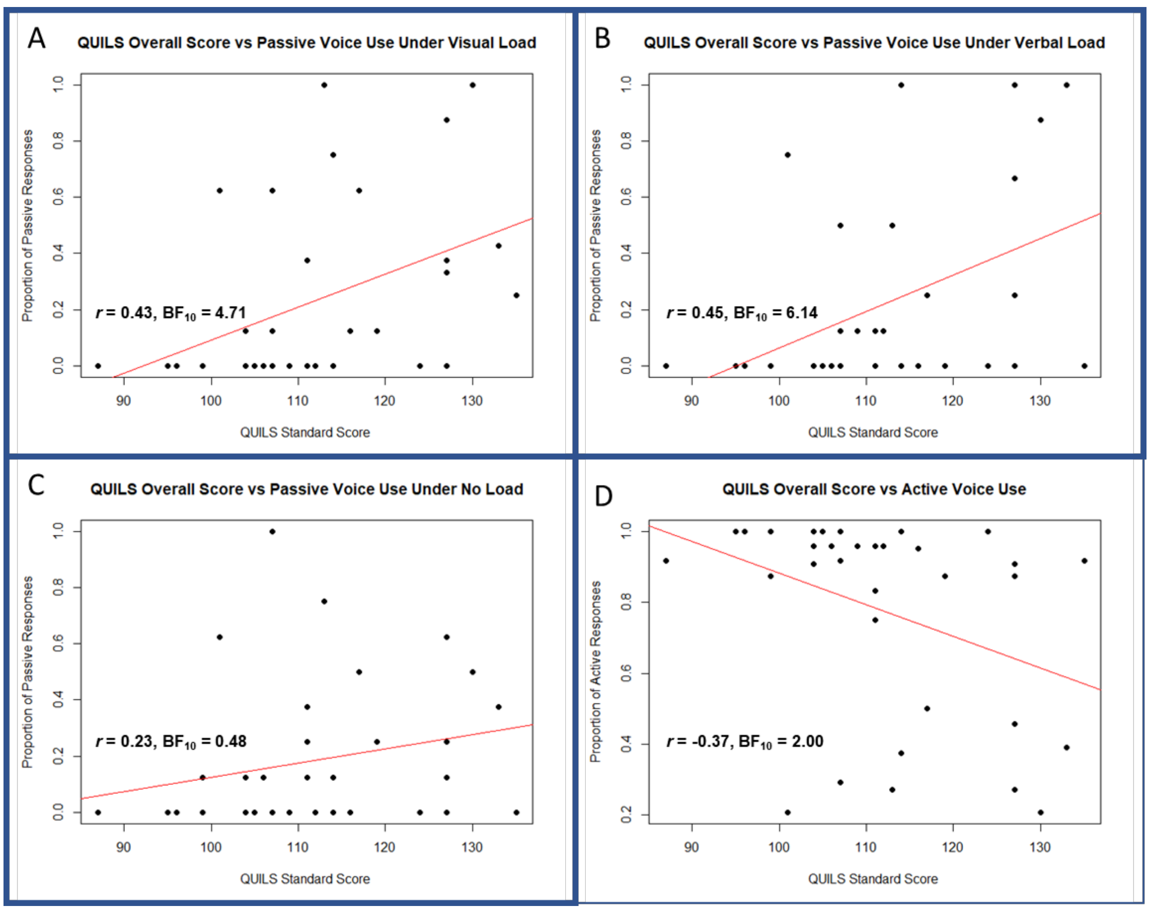


Figure 6. **A – C:** QUILS performance predicting passive voice use when: ***A***. required to hold a visual-spatial load; ***B***. required to hold an auditory-verbal load; ***C***. not required to hold any load. ***D***: QUILS performance against overall active voice use.

sample size and variance for these participants may have contributed to a lack of a true difference).

The three specific scores (i.e., vocabulary, syntax and processing) were not able to predict passive or active voice use on their own, except in one specific case. A higher standard score on the *syntax* portion of the QUILS predicted when participants spoke in the passive while under auditory-verbal load ($r = 0.44$, $BF_{10} = 5.38$). Overall QUILS performance was not able to predict when participants were more likely to make errors in the passive ($r = 0.21$, $BF_{10} = 0.422$) or active voice ($r = -0.07$, $BF_{10} = 0.23$). However, it did correlate with how well they performed on the load tasks ($r = 0.44$, $BF_{10} = 5.51$).

Possible covariates

Sex. One unusual characteristic about this group of participants was the slightly disproportionately high number of male participants. Because some literature suggests mild differences between male and female performance on language or verbal working memory (Eriksson et al., 2011; Bornstein et al., 2004), I assessed any differences in performance based on sex. All independent samples t-tests returned inconclusive or null results for differences on measures both in the experiment or auxiliary tests.

Age. Participants ranged in age from 4:0 to 5:9. I assessed if there were any measures that participant’s age alone could predict. Most results were inconclusive or null, except one. The older a participant was, the more accurate their responses on the auditory-verbal loads ($r = 0.49$, $BF_{10} = 16.69$). This relationship was not true for performance on the visual-spatial load task, however ($r = -0.10$, $BF_{10} = 0.24$). Age was not a significant factor for explaining overall performance on these tasks.

Discussion

In Study 3, I explored children's ability to speak in a relatively unfamiliar syntax using original constructions and tested how working memory loads affect their ability to do so. Next, I revisit each of the following key findings:

(1) Participants were able to construct the passive about 21% of all responses, but still overwhelmingly preferred to use the active voice.

(2) Working memory loads did not affect when participants chose to use the passive or active voice.

(3) Combined scores on vocabulary, syntax, and processing measures were able to predict when participants used the passive voice, and especially when they were under working memory load.

Able to produce the passive, but active voice preference maintained

One of the goals of Study 3 was to find if participants would be able to use the passive voice in their own sentence constructions, even after relatively little guidance. At the beginning of the experiment, each participant was provided with 3 model passive voice sentences that they repeated. Then, they were given 3 animations to practice describing different transitive actions using the passive voice. The experimenter provided feedback until they were able to produce or repeat a description in the correct way. This brief exposure seemed to be enough to encourage passive voice use for 21% of all responses, with 69.4% of participants being able to use the passive voice on their own at least once. However, most participants still used the active voice a majority of the time. This finding is unsurprising, as speakers of all ages will likely resort to what is most comfortable or familiar when under pressure to produce a sentence (Christianson et al., 2016).

As in the previous studies, each produced sentence was assessed for errors—syntactic or semantic elements that caused the sentence to deviate from the intended meaning. It is important to note that the nature of errors themselves changed between studies. In the first two studies, errors were easily flagged because each participant response was compared to the model passive voice sentence heard earlier in the experiment. In Study 3, because participants were tasked with creating brand new sentences about animations that had never been previously seen, errors could not be assessed in the same way. I assessed errors based on if they logically matched what was witnessed in each animation. The nouns and verbs a child used to describe the sentence would largely rely on their own lexicon, but the direction of the action should be somewhat universal. For example, one child may witness the animation of the horse pushing the cow and state, "The horse is shoving the cow," while another may choose to say, "The cow is being kicked back by the horsey." Although different verbs and word order were used, the perceived patient and agent are the same. However, if yet another child chose to describe the same animation as, "The cow is mad at that horse," the interpretation does not follow the same logic of (1) a transitive action taking place and (2) the horse being the agent of the situation.

The likelihood of participants describing the situation in a completely illogical manner was very rare and even sentences that were just slightly off-track from the intended action were rare. As in Study 2, the rates of errors within passive and active voice errors were similar. In this study, it could be that participants used whatever syntax seemed most achievable, leading to active voice use most of the time. Participants might have elected to only use the passive voice when there was a similar level of confidence (as with using the active voice) about producing a sentence that matched intended meaning. Thus, there were not a higher number of errors in the passive voice.

Study 3 demonstrates that participants are capable of using the passive voice in their own constructions, even if there remains a preference for the active voice. As noted, there is a possibility that specific sentences were more or less difficult than others for participants to describe in the passive voice. I provide rates of passive voice use for each sentence in Appendix D, but note that individual differences in experience with the words and actions used could affect a participant's likelihood of using the passive voice. Future studies could explore how adding more training throughout the experiment to increase confidence in using the passive, no matter the perceived difficulty, or adding an incentive to try to use the passive on more trials would affect production rates and production accuracy.

Working memory loads not effective

Of primary interest is the role of working memory in the ability to use the passive voice in original constructions. I predicted that participants would be able to use the passive voice the most often where there was no load, given that constructing new sentences in an unfamiliar syntax would likely be the most demanding task. However, there were no such patterns and working memory loads did not seem to affect passive voice use, at all. It could be the case that the passive voice task in this study was not demanding enough. Alternatively, it could be that constructing any original sentence about a transitive action, regardless of the syntax used, requires similar amounts of working memory. The working memory loads used in the study may not have been robust enough to cause differences in passive or active voice use if they both require high or similar amounts of working memory to produce. Some possible evidence for this heightened working memory tax is found in the performance on the load tasks. In Study 2, which only required recollection of passive voice descriptions, participants responded to load tasks with 45.5% accuracy. In Study 3, with the new requirement to construct new sentences, accuracy dropped to 37.5%, though the task and even load stimuli remained identical. It could be that participants experienced a dual-task cost—both forming sentences and remembering loads require working memory. Forgoing the load could possibly ensure a better chance of producing a logical sentence about each animation, thereby relieving the effect of the load altogether.

However, the decrease in accuracy on the load tasks from single-task trials to dual-task trials was nearly equal from Study 2 (38% decrease) to Study 3 (36.7% decrease). One caveat to these comparable decreases is that accuracy on practice trials for the visual-spatial loads was less accurate in Study 3 ($M = 0.61$) than in Study 2 ($M = 0.77$). If these accuracy values were more similar, a drop in performance may have been more evident. Alternatively, it could simply be the case that working memory is not the main predictor of children's ability to use the passive voice. I discuss evidence in favor of this hypothesis next.

Language abilities predict language skills

I collected measures of each child's vocabulary, syntax, and processing skills using the QUILS. Each child's scores on the subtests were combined to create a standard score that accounted for the child's age on the day of testing. Participants who scored higher on these measures were more likely to be able to use the passive voice, especially when under visual-spatial or auditory-verbal load. Better language skills did not predict their ability to use the passive voice when there was no load, however. These results may indicate two key points: (1) prior language experience still holds importance and (2) language and working memory skills may interact to predict ability to perform difficult production tasks.

Experts who support interactionist theories of language acquisition would likely predict the result that prior language knowledge and skills would be a significant contribution to the ability to produce sentences. If there is a possibility of a child being exposed to more passive voice constructions in the home or at school through conversation and reading, that child will be at advantage over children who rarely hear the construction. Previous studies have shown variations in the home according to SES or parental education can predict how much children hear constructions like the passive voice, which in turn predict their own mastery of the syntax (Corriveau et al., 2016; Huang et al., 2017). Additionally, children who are exposed to more book reading and turn-taking conversation may be more accustomed to the act of trying new and unfamiliar language. The task itself requires a level of approach and openness to using a syntax that is less familiar.

Performance on the QUILS language assessment was the best predictor for children's ability to use the passive voice, but specifically in cases when they were under load. It could be that prior language experience drives the ability to speak in challenging syntax, and is especially useful for cases in which the speaker is under working memory load. However, the lack of evidence for an effect of working memory load on passive voice use challenges this notion. Future work that focuses on increasing the use of the passive for all participants (whether through more training or incentives) could help explain these differences in passive voice use for more skilled speakers.

Chapter 6: General Discussion, Theoretical Assumptions, and Conclusions

In Adams & Cowan (2021) we asked how working memory loads affect 4- and 5-year-olds ability to recall passive voice descriptions of static images displaying a potential agent and patient. We found that participants were more likely to use the passive voice when under load and more likely to use the active voice when there was no load. We also found that participants produced more errors when trying to use the passive voice as compared to the active voice, and this seemed to be emphasized while under load. We concluded that participants must be using working memory to control for the quality of their speech by avoiding errors. The best way to avoid errors in speech is to speak in a syntax that is most familiar (in this case, the active voice). However, transforming a sentence from a relatively unfamiliar syntax to a more familiar one while maintaining intended meaning is a challenging task and requires working memory, hence the heightened likelihood of this occurring under no load.

In the two studies presented here, I aimed to build on these findings by exploring how children perform when they are given more encouragement to try to use the passive voice, both in recall and in their own original productions, through extra training and repetitions. In Study 2, I attempted to extend the findings of Adams & Cowan (2021; referred to here as Study 1) by adding more information about the passive voice, or the *special way* of talking, and giving participants a chance to practice recalling the passive sentences after a delay. Children were also required to repeat every probe sentence as they listened to each description of the transitive action images. I found that participants were significantly more likely to use the passive voice than those tested with the Study 1 method. Though passive voice use did not vary by working memory load type, active voice did (especially when extreme outliers were removed from the data). Participants were more likely to use the active voice when there was no load, replicating the finding from Study 1. Surprisingly, error rates did not increase from Study 1 to Study 2, despite the increased use of the passive voice. It could be that repeating each sentence after each probe trial served as a protective factor against increased errors, by means of verbal rehearsal. With increased passive voice use achieved, our next goal was to find if participants were able to produce their own passive voice sentences.

In Study 3, I created animations depicting transitive actions taking place between a patient and agent. Participants were introduced to the passive voice and also provided with practice trials for learning how to use the passive to describe the animations. Children then described 24 animations, some while under working memory load. Overall, participants were much more likely to use the active voice than the passive. However, roughly 21% of responses were passive voice descriptions. This achieves a slightly higher rate of passive voice use than in Huttenlocher et al. (2004) which used a priming method to indirectly encourage children to use the passive. In that study, children were able to use the passive in roughly 15% of their own image descriptions. The majority of participants in the present study were able to use the passive on their own at least once, demonstrating that even brief training can encourage most children to use the less common syntax. There were no clear effects of working memory load, and performance on the load did not reliably predict whether a child used the passive or active voice. However, combined standard scores on a separate language assessment of vocabulary, syntax, and processing skills predicted whether children were able to use the passive voice. This was especially true for cases in which the participant was under load.

What do these studies combine to tell us about the use of working memory for children's production of the passive voice? I suggest that results begin to address some of the following issues:

(1) The role of working memory in recalling language structures versus original, directed production

(2) The role of working memory for avoiding errors and producing intended meaning

(1) The role of previous language skills in predicting new language task performance

After discussing how the present results relate to each of these issues, I discuss what this could mean for the child speaking in their natural environment and conclude with possible future directions for continuing to answer remaining questions about these processes.

Table 1. Cross-study comparisons

Measure	Study 1		Study 2		Study 3		Average across studies
	M	SE	**M**	SE	**M**	SE	
Active voice use	**0.26**	0.04	**0.10**	0.02	**0.78**	0.05	0.38
Passive voice use	**0.65**	0.05	**0.81**	0.02	**0.21**	0.05	0.56
Error rates							
Active voice	**0.08**	0.03	**0.15**	0.06	**0.09**	0.02	0.11
Passive voice	**0.17**	0.04	**0.10**	0.02	**0.15**	0.05	0.14
Verbal load performance							
Single-task	**NA**	NA	**0.93**	0.03	**0.88**	0.05	0.91
Dual-task	**0.35**	0.04	**0.42**	0.03	**0.38**	0.03	0.38
Visual load performance							
Single-task	**NA**	NA	**0.77**	0.04	**0.61**	0.06	0.69
Dual-task	**0.44**	0.03	**0.49**	0.04	**0.37**	0.04	0.43

Table 1 Note. (1) Error rate refers to the proportion of errors out of all errors of the respective syntax. (2) Single-task data for Study 1 is not available as these trials were added to Study 2 design.

Working memory for recall versus original production

The contrast in passive voice use between Studies 2 and 3 begin to tell a story about the starkly different tasks of recalling a sentence versus constructing a new sentence, and how working memory contributes to each. When children were asked to recall passive voice structures, they were able to do so about 80% of all responses, but when children were asked to use the passive voice in their own new sentences, this rate dropped to around 21% (cross-study comparisons of selected measures are provided in Table 1). Results also showed that working memory loads seemed to be more impactful on the recall of sentences than original production. At first glance, this appears to be contradictory. It might be assumed that recalling a previously heard sentence should not require as much working memory power as creating a brand-new sentence about events that had not been previously witnessed. I suggest that even in light of the results of these studies, this assumption can still hold true when considering how these processes might work in the embedded processes model of working memory.

In Study 2, when participants are required to recall passive voice descriptions that have been previously heard, they are first presented with the most useful cue of the image itself. Upon viewing the image, all the relevant elements (and perhaps some related but irrelevant ones) should be activated in long-term memory. Then, central executive processes should be able to pull the correct words into the correct order in the focus of attention to be prepared for production. Because the sentence is already a part of activated long-term memory due to rapid learning during the repetitions phase, it is possible that not as much attention is needed to deal with the sentence as it might be when the sentence was only heard (like in Study 1). Central executive processes may also be able to perform a quality check on the sentence before production to ensure it matches intended meaning. However, because of the child’s lack of experience with the passive, the best way to ensure the sentence matches the intended message is to transform it into the active voice. This is easiest to do when the focus of attention has enough space and central executive processes are not over-taxed. If, however, the focus of attention is occupied by other information (e.g., the working memory loads), there may not be enough resources left over to handle transforming the sentence and checking the order against intended meaning. The least burdensome action may be to simply attempt to restate the same sentence in the passive voice. Without the necessary working memory to check the quality of the sentence, producing errors is still a possibility, given the similar number of errors within Studies 1 and 2. However, because of the sentence repetition requirement in Study 2, the sentence may be more strongly represented in activated long-term memory, preventing higher rates of errors in the passive voice as compared to the active voice.

In Study 3, the task demand shifted considerably to describing animations that had never been viewed while also attempting to use the passive voice after a brief introduction. Impressively, the majority of children were able to use the passive in at least one of their own responses, but still preferred to use the active in most of their responses. Working memory loads did not impact when they used the passive voice. In the Study 3 method, when a child is presented with an animation, the first task is to discern what is occurring. This may require some level of working memory as the participant must keep track of the characters, who is doing what and to whom, and any other relevant details about the animation. Next, after deciphering the action taking place, the child must find the correct words to assign to the correct characters. Unlike Study 2 in which participants were already equipped with correct words from the probe sentence, now participants were required to derive the correct label themselves from their own lexicon. This would require some searching in long-term memory, then the correct words (and related words) would be activated in long-term memory. Next, central executive processes would be tasked with putting the words into the correct order and expressing intended meaning according to the child's interpretation. Now, there is no default passive sentence to rely on if working memory is occupied (as in Study 2). Participants' only options are to formulate the new sentence using either the familiar active voice or the passive voice, both of which require a high amount of working memory.

It could be that this task demand, no matter the syntax used, was so high that the working memory load itself became secondary in priority and the load may have been sacrificed to achieve the language task. After all, participants could guess or indicate complete forgetting on the load task if necessary, but it is not possible to "guess" on a description task. This partial or complete forgoing of the load portion of the test trials might have been evident in accuracy rates on the load task. However, there was inconclusive evidence for a true decrease in load performance from Study 2 to Study 3, however, so this possibility remains uncertain. In sum, it could be that working memory is certainly needed for producing original sentences, perhaps even more so than recalling sentences. When the language task is prioritized over the load task, however, the load becomes less impactful to the point of not causing any effects. To solve this, a future study could emphasize the importance of performing better on the load task by providing trial-by-trial feedback or increasing incentive to remember the digits or token locations.

An alternative explanation may be supported by theories about syntactic processing and working memory proposed by Waters & Caplan (1999). In Waters et al. (2003), college-aged participants were given a syntactic processing task that involved judging the plausibility of sentences that ranged in syntactic complexity. In some conditions, participants were required to hold a concurrent load (2 digits) while processing the sentences. Although listening time increased for more complex sentences and concurrent loads also increased listening time, it did not increase from simple to complex sentences while under load. If it were the case that a general working memory resource is needed to both process the sentences and hold digits, it would be expected that processing of more complex sentences would become more difficult under load as compared to simple sentences. This was not the case and the authors suggested that syntactic processing occurs in a specialized working memory that is not affected by non-syntactic information.

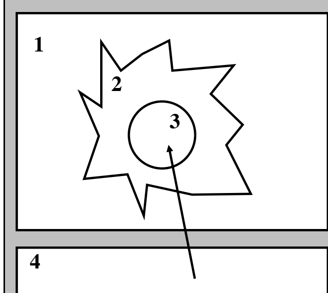
Waters and Caplan's theory could explain the results of Study 3, but cannot fully explain results from the previous two studies. In Study 3, it could be that children were trying to hold the load but were constructing the sentences using this specialized syntactic working memory, so working memory loads were not effective. For Study 2, the task may have required less syntactic processing and more rote memory, which pulls on similar working memory power as the loads. However, participants still preferred to use the active voice under no load, suggesting that transforming the sentence required some amount of working memory. As with other remaining questions, one solution to uncovering how working memory is used for various tasks may lie in emphasizing the importance of maintaining and correctly recalling the load itself. It will also be important to disentangle how much of an additional load the sentence construction task is in Study 3, as these actions should have been less ambiguous than the static images which depicted no clear verb.

Working memory for avoiding errors in speech

Results from Study 2 mirrored one of the primary effects from Study 1 in that participants were more likely to use the active voice when there was no additional load to hold (especially when outliers were handled). I proposed that this paraphrasing may actually be a way to ensure that the intended message is being portrayed and that desire still seems to hold in Study 2. For Study 3, errors became more nuanced because there was no model to compare the response sentence to. It is also possible that by defaulting to the active voice, participants were attempting to avoid labeling and role assignment errors that would have occurred if using the passive. However, there was no evidence for a difference in proportion of errors within each specific syntax type used. Participants produced errors at similar rates whether they were trying to use the passive or active voice. It could be that, when constructing their own sentences, participants varied syntax by the likelihood of stating the sentence correctly. The passive voice sentences that were produced may have occurred in cases in which participants were trying to follow instructions but also if they felt comfortable enough to use the structure. Otherwise, using the active voice may have been the safest option for avoiding contradicting intended meaning.

Despite the low rates of errors that occurred in Studies 2 and 3, it is useful to assess what *types* of errors were occurring and how these errors can be explained in light of the embedded processes model as well as Bock and Levelt's grammatical encoding model. The different stages of Bock and Levelt's model may be handled by different parts of the embedded processes model depending on the language task at hand. For example, part of the grammatical encoding process is lexical assignment. When recalling previously heard sentences, because all lexical items have already been provided, this task may be handled more in activated long-term memory as opposed to the focus of attention. In turn, when constructing a new sentence, the focus of attention may be more involved in retrieving items from long-term memory while viewing the animation. At the same time, related lexical items are activated in long-term memory. The function assignment step of Bock and Levelt's model may also take place in different parts of the embedded processes model depending on the task. During recall, the roles of each noun should be retrievable from long-term memory and easily activated when the cue is presented. For original production, the functions are assigned by the participant as they

Figure 7. Role of embedded processes in two language tasks



Part of model	Recalling passive sentence	Constructing new passive sentence
1. Long term memory	Storage of previously heard sentences	Provides lexicon and previous syntactic knowledge
2. Activated long-term memory	Upon cue, elements of previously heard sentences activated including roles of nouns	As labels are retrieved, appropriate words are activated
3. Focus of attention	Model sentence and intended message are actively held in mind	-Potential lexical items held in mind -Animation is interpreted for meaning
4. Central executive processes	Sentence put into desired order, and compared to model sentence for intended message	-Nouns are assigned functional roles -Sentence is evaluated against intended message and put into desired order

are being discerned from the animation. In Figure 7, I provide a schema for possible contributions of the various subparts of the embedded processes model in the grammatical encoding process.

The proposed tasks of each model can be used to predict where errors are occurring. If a child is recalling a sentence and produces a lexical assignment error, perhaps the error is occurring because of a disruption at the level of activated long-term memory due to temporal decay (since the model sentence was last heard). If a lexical assignment error takes place during original production of a sentence, the error could take place within long-term memory because of a lack of the correct word or because the word representation is not strong enough. Alternatively, the disruption could take place within the focus of attention due to overtaxed central executive processes. Regardless, because all these processes are proposed to be entangled in nature, it is difficult to definitively conclude that disruptions have taken place in one process and not another, but these models may combine to begin to help demonstrate the complexity of each language task and how these processes work together to produce a sentence.

Language skills still matter

One of the goals of the present work was to illustrate the importance of cognitive processes, namely working memory, for the ability to produce language. I began with an understanding that there is a strong body of literature demonstrating that other factors, such as language experience and knowledge, often drive children's language development. A portion of results from Study 3 align with this view. When scores on vocabulary, syntax, and processing measures on the QUILS were combined, it was the best predictor of which participants were able to use the passive voice. It was especially useful for predicting when participants were able to use the passive voice while also attempting to hold an additional memory load. These results fit in well with secondary findings in Study 1 in which the Expressive Vocabulary Test was a predictor of children's ability to recall the passive voice, and especially without error.

These results combine to demonstrate how previous language skills serve as a critical foundation for tackling new language skills. Not even the age of participants was able to reliably predict performance on the language task as did standardized performance on the QUILS. I consider two possibilities for why this relationship exists: (1) participants who have experience with the vocabulary of the sentence or the passive voice will ultimately be better at using it during the task and (2) experience with approaching new language tasks serves as good practice. First, the vocabulary children approached the experiment task with was certainly an individual difference that should be carefully considered. If evaluating vocabulary sizes and depth in terms of the embedded processes model, a participant's previous vocabulary and syntax knowledge is a part of their long-term memory. The more established a word or sentence form is in long-term memory, the easier and more accurately it is recalled at appropriate times. Thus, for a child who has known the word "giraffe" for 2 years, the task of discussing a giraffe popping a balloon might be less effortful than for a child who just learned the word "giraffe" 2 weeks previously. Taking some of the effort of word retrieval away, the more experienced speaker may have more working memory left to attempt to use the passive voice, thus giving one possible explanation the correlation between language scores and passive voice use.

Second, participants who are used to "flexing the muscle" of language learning more often through conversation, reading, or spontaneous self-talk may be better at approaching challenging language tasks such as the one in Study 3. These children will ultimately score better on tests such as the QUILS (which measures how well children acquire new language) and subsequently be able to use the passive voice more in the experiment. In either case, experience and knowledge seem vital for being able to use the passive voice in original constructions.

Implications for everyday language

In everyday conversation, most children are not likely to be asked to recall passive voice sentences or describe strange events only using the passive voice. However, the essence of language acquisition is the process of learning, practicing, and correctly using new words and syntax in comprehendible ways. The results of these studies may give us a hint about what is necessary for these tasks. Working memory appears to be helpful for portraying intended meaning, which could mean paraphrasing what has been heard into a more familiar form. This can be witnessed in conversations with children in which they recast a sentence from a more mature speaker into their own preferred form (e.g., “You went to the park?” “Yes, I goed to the park.”). In these cases, it could be that the child’s intended meaning matches that of their conversational partner, but they are not yet ready or able to use the more grammatically correct version of the sentence. Though this highlights linguistic limitations, the present studies may suggest that this actually requires working memory to successfully do.

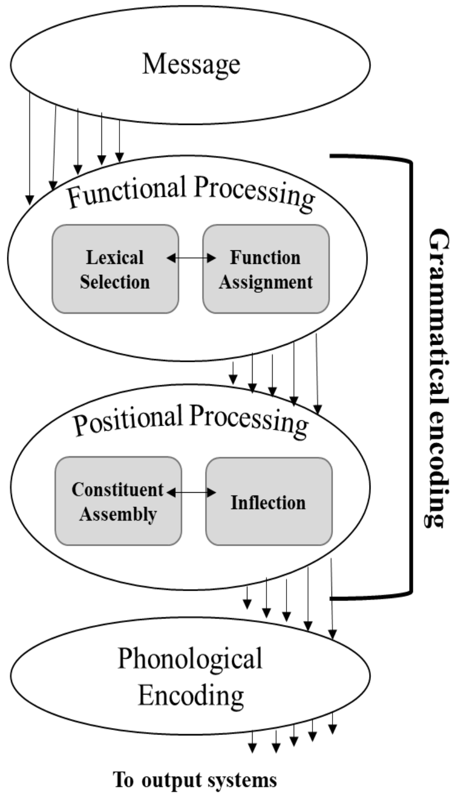
These results also suggest that previous knowledge helps children use new or relatively unfamiliar language more proficiently. This gives even more reason for caregivers and educators to expose children to a wide variety of language, as it can help them to approach difficult language tasks later on. Part of results also appeared to suggest that this prior language skill was especially useful for using the passive voice when they were trying to hold an additional load. It could be that language skills serve as a protective factor when working memory is taxed, although these claims are limited by the lack of an overall effect of working memory load on passive voice use in Study 3. Altogether, these studies serve as testament to the difference between reproducing a sentence heard before and constructing a new sentence and shed light on how working memory may have different roles for each task. The best way to ensure children are given the most conducive environment for learning new language is to limit distractions, decrease memory load when able, and consider previous language knowledge.

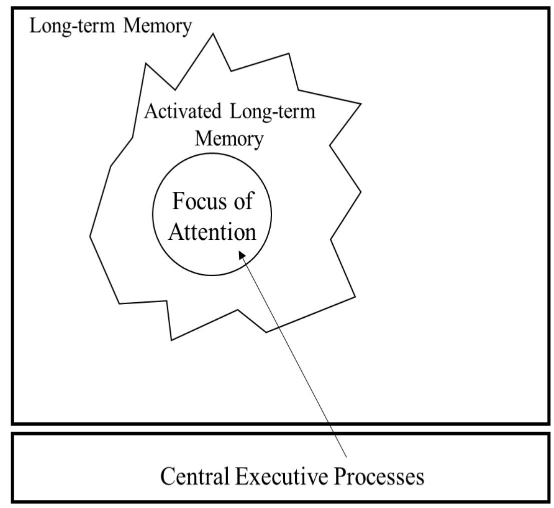
Future directions

These studies begin to answer questions about the relationship between working memory and language, but implications of our findings are limited by a few factors. For one, while the age range studied in these samples are appropriate for the task and research questions, language development certainly extends beyond the 4- to 5-year-old range. Future studies could assess how older children perform on the language task to continue to capture the developmental trajectory of the skills being assessed. Second, although I was able to obtain digit span scores for the majority of participants and also assess performance on the load task, a true estimate of each child's individual working memory capacity was limited. Future studies could add another measure of working memory capacity to examine how capacity is able to predict performance on all the measures. Finally, as noted, it is possible that participants choose to forgo the loads in cases of high task demand, which may negate the purpose of the loads. This could be manipulated in future studies to emphasize the importance of accuracy on the load task, which would allow for a better examination of how constraining working memory affects the language tasks.

The two studies presented here in addition to the work of Adams & Cowan (2021) serve as an example of how researchers can begin to study the connection of working memory and language through experimental means. As more evidence is added, it becomes possible to make more causally based claims about the connections between these two processes if the causal relationship does exist. Discovering more about the role of working memory in language development may open up doors to better assisting children as they grow and learn how to tackle more complex language tasks with age.

Appendix A. Models for theory background.





Left: Re-creation of Bock and Levelt (1994) model of grammatical encoding.
Right: Simplified depiction of Cowan (1988, 1999) Embedded-processes model.

Appendix B. Effects on error (or correct sentence) rates and specific error types (next pages)

Study 2 Error Effects

Response Type Category	**Overall rate (M, SE)**	**Load Type Effect (BF_{incl})**	**Reversibility (BF_{incl})**	**Interaction Term (BF_{incl})**
Correct Passive	0.72, 0.03	0.047	0.127	0.011
Incorrect Passive	0.08, 0.02	0.397	1.026	0.127
Correct Active	0.09, 0.02	0.514	3.844*	0.118
Incorrect Active	0.01, 0.00	0.268	0.389	0.242

*This BF_{incl} is reflective of the effect of keeping reversibility in the model as a factor. The BF_{10} included in the main text is the result of an independent t-test.

Study 2 Error Types and Their Rate of Occurrence

Type of error	**Description**	**Example**	**Study 2 Rates**
Direct reversal	□ Switches patient and agent roles (i.e., actor becomes acted upon AND vice versa) □ Maintains verb (including accepted alternatives, e.g., push = shove)	*The boat was carried by the man* becomes *The man was carried by the boat*	**62.02%**
Half loss	□ Alters one agent (other agent remains in correct role) □ Maintains verb	*The pig was touched by the turkey* becomes *The pig was touched by the zebra*	**13.92%**
Reversal without verb	□ Switches patient and agent (i.e., actor becomes acted upon AND vice versa) □ Verb changes meaning	*The car was pulled by the horse* becomes *The horse was carried by the car*	**12.66%**
Complete loss of semantics	□ Alters one or both subjects of sentence AND □ Omits or alters the verb	*The television was turned on by the girl* becomes *The girl was screamed by the me.*	**6.32%**
Half reversal	□ Actor becomes acted upon OR vice versa (other agent is altered) □ Verb maintained	*The girl was pulled by the dog* becomes *The dog was pulled by the bottle.*	**5.06%**
Verb only	□ Maintains agent and patient roles □ Changes verb meaning	*The brother was pushed by the sister* becomes *The brother was kicked by the sister.*	**0%**

Note: These error types, descriptions, and examples are adapted from Adams and Cowan (2021; Study 1).

Study 3 Error Types and Their Rate of Occurrence

Type of error	**Description**	**Example** *Select example from data* Intended sentence	**Study 3 Rates**
Missing agent	Participant only describes one noun without eluding to other	*"Mom sat down."* The girl was hugged by the mom.	**43.08%**
Labeling error	Participant labels patient or agent incorrectly	*"The pumpkin is being kicked by the horse."* The pumpkin was kicked by the horse.	**23.08%**
Truncation	Truncating the sentence by leaving out a noun (but the omitted noun is still implied)	*"Cutting a sandwich."* The sandwich was cut by the boy.	**20.00%**
Direct reversal	Participant switches patient and agent in the action	*"The alligator got picked up by the bird."* The alligator was jumped on by the bird.	**6.15%**
Naming	Participant simply labels or names the nouns without mentioning an action	*"The duck, the panda bear."* The panda was jumped on by the bird.	**4.62%**
No Action	Participant describes scenario, but does not include a transitive action.	*"The alligator doesn't want a bird on his back."* The alligator was jumped on by the bird.	**3.08%**

Appendix C. Animations and Sentence Stimuli from Study 3

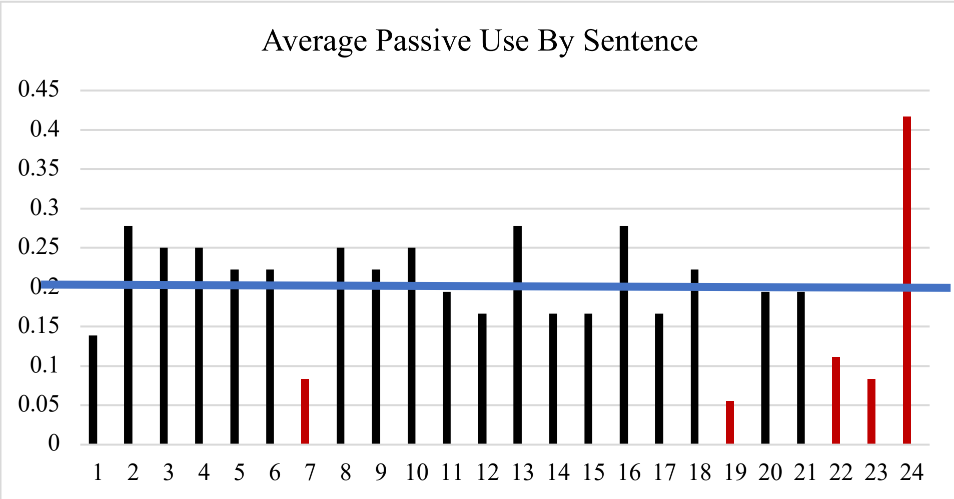
https://osf.io/8unfs/?view_only=3289c92dac164bf28becb18ee1dfe2c4
Note: Because all participants were asked to devise their own passive voice constructions about each of these animations, these sentence stimuli only serve as a guide for the intended *meaning of the animation. These sentences are presented in order of the animations on the OSF site.*

*used for training trials

OSF site label	**Intended action**	**Side of agent**
Bird-alligator	The alligator was jumped on by the bird.	Left
Bird-panda	The panda was jumped on by the bird.	Left
Boy-book	The book was dropped by the boy.	Right
Boy-cookie	The cookie was eaten by the boy.	Left
Boy-sandwich	The sandwich was cut by the boy.	Left
Dad-bear*	The bear was picked up by the dad.	Right
Dog-cat	The cat was chased by the dog.	Right
Dog-door	The door was closed (shut) by the dog.	Right
Dog-girl	The girl was pulled by the dog	Left
Duck-chicken*	The chicken was splashed by the duck.	Left
Elephant-ball*	The ball was kicked by the elephant	Left
Firefighter-house	The house was sprayed by the firefighter.	Right
Frog-bug*	The bug was eaten by the frog.	Right
Giraffe-balloon	The balloon was popped by the giraffe.	Left
Girl-boy	The boy was waved to (at) by the girl.	Right
Girl-flower	The flower was watered by the girl.	Right
Girl-TV	The TV was turned off by the girl.	Right
Goat-duck	The duck was chased by the goat	Right
Horse-car	The car was pulled by the horse.	Right
Horse-cow	The cow was pushed by the horse	Right
Lion-bear	The bear was pushed by the lion	Left
Man-boat	The boat was carried by the man	(Carried from right)
Mom-girl	The girl was hugged by the mom.	Right
Mom-soup*	The soup was stirred by the mom.	Right
Monkey-banana	The banana was eaten by the monkey.	Left (holding)
Monkey-bottle	The bottle was thrown by the monkey.	Right

Sun-ice-cream*	The ice cream was melted by the sun.	Right
Teacher-candles	The candles (cake) were blown out by the teacher.	Right (holding)
Turkey-pig	The pig was bumped by the turkey.	Left
Zebra-pumpkin	The pumpkin was kicked by the zebra.	Left

Appendix D. Item Level Analysis of Passive Voice Use



#	Sentence	#	Sentence
1	The alligator was jumped on by the bird.	**13**	The flower was watered by the girl.
2	The banana was eaten by the monkey.	**14**	The balloon was popped by the giraffe.
3	The boat was carried by the man	**15**	The girl was pulled by the dog
4	The book was dropped by the boy.	**16**	The duck was chased by the goat
5	The bottle was thrown by the monkey.	**17**	The girl was hugged by the mom.
6	The candles (cake) were blown out by the teacher.	**18**	The bear was pushed by the lion
7	The car was pulled by the horse.	**19**	The panda was jumped on by the bird.
8	The cookie was eaten by the boy.	**20**	The sandwich was cut by the boy.
9	The cow was pushed by the horse	**21**	The pig was bumped by the turkey.
10	The cat was chased by the dog.	**22**	The TV was turned off by the girl.
11	The door was closed (shut) by the dog.	**23**	The boy was waved to (at) by the girl.
12	The house was sprayed by the firefighter.	**24**	The pumpkin was kicked by the zebra.

Bars in red are those that fell 1 standard deviation above or below the mean of $M = 0.21$.

***order of sentences is due to programming labels**

References

Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences?: A permutation analysis. *PLOS ONE*, *12*(10), e0186129. https://doi.org/10.1371/journal.pone.0186129

Adams, A.-M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research*, *38*(2), 403–414. https://doi.org/10.1044/jshr.3802.403

Adams, E. J., & Cowan, N. (2021). The girl was watered by the flower: Effects of working memory loads on syntactic production in young children. *Journal of Cognition and Development*, *22*(1), 125–148. https://doi.org/10.1080/15248372.2020.1844710

Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools*, *49*(3), 340–355. https://doi.org/10.1044/2018_LSHSS-17-0114

Alloway, T. P., & Gathercole, S. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology*, *17*(2), 207–220. https://doi.org/10.1080/09541440440000005

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, *47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure: Overgeneralization in child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 47–62. https://doi.org/10.1002/wcs.1207

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*(2), 234–238. https://doi.org/10.1016/S0959-4388(98)80145-1

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Baldie, B. J. (1976). The acquisition of the passive voice. *Journal of Child Language*, *3*(3), 331–348. https://doi.org/10.1017/S0305000900007224

Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist*, *31*(2), 133–140. https://doi.org/10.1207/s15326985ep3102_5

Bidgood, A., Pine, J. M., Rowland, C. F., & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained: Evidence from children's and adults' comprehension and production/priming of the English passive. *Cognitive Science*, *44*(9). https://doi.org/10.1111/cogs.12892

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development*, *17*(1), 91–107. https://doi.org/10.1177/016502549401700106

Bloom, L. (1976/2013). *One word at a time: The use of single word utterances before syntax* (Vol. 154). Walter de Gruyter.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode.

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2003). Neural development of selective attention and response inhibition. *NeuroImage*, *20*(2), 737–751. https://doi.org/10.1016/S1053-8119(03)00404-X

Brooks, P. J., & Tomasello, M. (1999). Young children learn to produce passives with nonce verbs. *Developmental Psychology*, *35*(1), 29–44. https://doi.org/10.1037/0012-1649.35.1.29

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

Carey, S., & Bartlett, E. (1978). *Acquiring a single new word*.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629–647. https://doi.org/10.1002/wcs.63

Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1–61.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Corriveau, K. H., Kurkul, K., & Arunachalam, S. (2016). Preschoolers' preference for syntactic complexity varies by socioeconomic status. *Child Development*, *87*(5), 1529–1537. https://doi.org/10.1111/cdev.12553

Cotton, K., & Ricker, T. J. (2021). Working memory consolidation improves long-term memory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 208–219. https://doi.org/10.1037/xlm0000954

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, *145*(8), 822–847. https://doi.org/10.1037/bul0000199

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations: Differences in working memory capacity. *Developmental Science*, *14*(5), 1066–1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x

Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, *18*(1), 132–145. https://doi.org/10.1111/desc.12197

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*, *143*(5), 1806–1836. https://doi.org/10.1037/a0036814

Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*(6), 604–615. https://doi.org/10.1016/j.langsci.2005.11.014

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Elliot, E. M., Morey, C.C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E., … Voracek, M. (2021; *in press*). Multi-lab direct replication of Flavell, Beach and Chinsky (1966): Spontaneous verbal 2 rehearsal in a memory task as a function of age. *Advances in Methods and Practices in Psychological Science*.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, *37*(2), 283. https://doi.org/10.2307/1126804

Fox, D., & Grodzinsky, Y. (1998). Children's passive: A view from the *by* -phrase. *Linguistic Inquiry*, *29*(2), 311–332. https://doi.org/10.1162/002438998553761

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*(1), 136–158. https://doi.org/10.1016/j.jml.2004.03.008

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *26*(2), 204–216. https://doi.org/10.1037/0012-1649.26.2.204

Golinkoff, R. M. (1981). The case for semantic relations: Evidence from the verbal and nonverbal domains. *Journal of Child Language*, *8*(2), 413–437. https://doi.org/10.1017/S0305000900003275

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition*, *36*(3), 227–254. https://doi.org/10.1016/0010-0277(90)90058-R

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*(6), 236–242. https://doi.org/10.1016/j.tics.2007.04.001

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, *44*(4), 339–387. https://doi.org/10.1006/cogp.2001.0769

Hirsch, C., & Wexler, K. (2006). Children's passives and their resulting interpretation. *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition*, *4*, 125–136.

Hoff-Ginsberg, E. (1990). Maternal speech and the child's development of syntax: A further look. *Journal of Child Language*, *17*(1), 85–99. https://doi.org/10.1017/S0305000900013118

Huang, Y. T., Leech, K., & Rowe, M. L. (2017). Exploring socioeconomic differences in syntactic development through the lens of real-time processing. *Cognition*, *159*, 61–75. https://doi.org/10.1016/j.cognition.2016.11.004

Huttenlocher, J. (1998). Language input and language growth. *Preventive Medicine*, *27*(2), 195–199. https://doi.org/10.1006/pmed.1998.0301

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language*, *50*(2), 182–195. https://doi.org/10.1016/j.jml.2003.09.003

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. H., Gustafsson, J., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. https://doi.org/10.1111/desc.12202

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1–10. https://doi.org/10.1162/089892902317205276

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, *59*(2), 200–222. https://doi.org/10.1016/j.jml.2008.04.004

Lidz, J., & Gleitman, L. R. (2004). Yes, we still need Universal Grammar. *Cognition*, *94*(1), 85–93. https://doi.org/10.1016/j.cognition.2004.03.004

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition*, *89*(3), 295–303. https://doi.org/10.1016/S0010-0277(03)00116-1

Logie, R., Camos, V., & Cowan, N. (Eds.). (2020). *Working memory: The state of the science*. Oxford University Press.

Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during covid-19. *Trends in Cognitive Sciences*, *24*(8), 583–584. https://doi.org/10.1016/j.tics.2020.05.003

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), i. https://doi.org/10.2307/1166115

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992–1009. https://doi.org/10.1037/bul0000108

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*(1), 1–11. https://doi.org/10.1017/S0305000900001276

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, *3*(1), 285–308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families: Maternal correlates of growth in toddler vocabulary production. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology*, *115*, 101238. https://doi.org/10.1016/j.cogpsych.2019.101238

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure?: Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders*, *50*(1), 106–118. https://doi.org/10.1111/1460-6984.12126

Ricker, T. J., Spiegel, L. R., & Cowan, N. (2014). Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1510–1523. https://doi.org/10.1037/xlm0000018

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211–227. https://doi.org/10.1037/0096-3445.126.3.211

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Thomason, M. E., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G. H., & Gabrieli, J. D. E. (2009). Development of spatial and verbal working memory capacity in the human brain. *Journal of Cognitive Neuroscience*, *21*(2), 316–332. https://doi.org/10.1162/jocn.2008.21028

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *Journal of Neuroscience*, *34*(5), 1592–1598. https://doi.org/10.1523/JNEUROSCI.0842-13.2014

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60. https://doi.org/10.1037/h0044616

Vanormelingen, L., Faes, J., & Gillis, S. (2020). Language development in children from different SES backgrounds: Babbling onset and consonant characteristics. *Dutch Journal of Applied Linguistics*, *9*(1–2), 132–161. https://doi.org/10.1075/dujal.19032.van

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

VITA

Eryn Adams was born March 24, 1994 in Jonesboro, Arkansas where she was raised. She attended Nettleton elementary, intermediate, and high school from which she graduated in 2012. She majored in Psychology and minored in Spanish at the University of Arkansas where she received her Bachelor of Arts in May 2016. She completed her Master of Arts under the guidance of Dr. Nelson Cowan at the University of Missouri in 2018 studying working memory and language development. She also completed her doctorate degree under the same advisory in August 2021.