Chapter 6 – General Discussion

Older adults often experience language production declines, such as using syntactically simpler speech or having increased word retrieval failures (Burke & Shafto, 2004; Kemper, Thomas & Marquis, 2001). On the other hand, language comprehension abilities remain relatively stable across the adult lifespan. However, previous research suggests that language prediction, which occurs during comprehension, also shows age-related decline (Federmeier, & Kutas, 2005; Kutas et al., 2010). Additionally, language production theories suggest that language prediction and production may be related processes, with Dell and Chang (2014) even proposing that language prediction *is* language production. However, research on the relationship between language prediction and language production, has been sparse. Therefore, I conducted a series of four self-paced reading studies in younger and older adults to further our understanding of the language prediction–production relationship and how this relationship might differ with age.

The specific study hypotheses can be reviewed in Table 2; however, to summarize, I hypothesized that overall, younger adults would demonstrate better language production and language prediction abilities compared to their older adult counterparts. Additionally, previous research indicates that older adults may engage more in language prediction under supportive circumstances, like when provided with richer semantic context. Therefore, I hypothesized that with additional context, older adults would show greater behavioral performance differences than the younger adults. Finally, I hypothesized that in the self-paced tasks with an online production component, the expectation to produce would encourage participants to engage in prediction.

Language Production Performance

Across the four studies, the younger adults largely had higher language production composite scores than the older adults, apart from study one, which supports the hypothesis that younger adults would demonstrate better language production abilities. Given the volume of literature finding evidence for age-related declines in production, the absence of a significant age-group difference in language production composite scores in study one was surprising (Burke, MacKay, & James, 2000; Burke, MacKay, Worthley, & Wade, 1991; Burke & Shafto, 2004; Burke & Shafto, 2008; Kemper, Thompson, & Marquis, 2001). When examining the data from study one, there are a few possibilities that may underlie this non-significant finding, such as the sensitivity of typed tasks, issues of data quality, and the specific population of participants recruited.

One limitation in conducting these studies online was not obtaining overt language production data. At the time the studies were launched, the microphone component in PsychoPy was not compatible with online use. Therefore, all language production measures had to be typed. While typing is still language production, it is inherently different from measures containing overt articulation. Many of the language production differences we typically observe, such as word retrieval failures, more disfluencies, or a slower speaking rate, are less detectable during typed tasks. Additionally, typing affords participants more time and allows them to be more intentional in their responses. Therefore, it is not surprising that during study one, one of the only significant age-group differences was picture naming response time, since older adults generally exhibit declines in response speed. Future studies examining the relationship between language prediction and language production should use overt measures of articulation, which allows researchers to observe and measure more components of language production difficulties, including word retrieval failures, production disfluencies, or speaking rate. However, considering

that there were significant differences in studies two through four, typed production is still a valuable tool for researchers as it can be used to collect data online from a broader population. While typing alone cannot explain the non-significant finding in study one, it may be a contributing factor. Therefore, typing should be used in conjunction with measures of overt articulation to have a fuller picture of production abilities.

When considering the absence of language production age-group differences in study one, the younger adults may have underperformed. The younger adult sample in study one was recruited from the psychology department subject pool. Because students are required to participate in research for course credit, there is concern that they do not put much effort into the studies they complete. To explore this, the language production scores for study one younger adults were compared to the younger adult production scores in studies two through four, in which many younger adults were recruited using Prolific. Overall, the younger adults in study one did not perform significantly worse on the language production tasks compared to the younger adults in the other studies. Since younger adults performed comparably across studies, I must turn to the older adults' performance in study one.

To explain the absence of language production differences in study one, the older adults may have performed comparably to younger adults. I compared older adult language production performance in study one to older adult language production performance across the other three studies. Results indicated that older adults in study one had significantly higher production scores compared to the older adults in studies two through four, suggesting the older adults in study one performed like younger adults on the language production tasks.

This finding speaks to a larger concern in aging research. Oftentimes, the older adults we recruit are aging exceptionally well compared to older adults in the general population. The

inclusionary criteria and participation requirements (e.g., access to a computer in the current study, or access to transportation for in-person studies) for aging studies favor recruiting relatively healthy, highly educated, and high socioeconomic status older adults. Consequently, these older adults do not always exhibit the same degree of cognitive decline that we might observe in a broader sample. Moreover, the aging literature theorizes that although individual variability increases across the lifespan in the general population, the older adults who participate in studies are a more homogenous group of individuals than their younger adult counterparts (Cabeza et al., 2018). Consistent with this, many of the older adults on Prolific have participated in hundreds of different studies. As a result, they may have completed similar tasks from other research groups or may be a top performer in their age cohort. However, one advantage of using online recruitment tools like Prolific is the ability to prescreen participants. This allows researchers to target specific qualities in participants (e.g., geographic location, education level, occupation, etc.). While not done in the current study, the increase in online data collection, and the use of tools like Prolific, may be a valuable asset to aging researchers in recruiting a wider variety of older adults.

Language Prediction Performance

In studies one through four, participants completed a version of a self-paced reading task to study language prediction. Analyses were run across all four studies to determine if there was a main effect of sentence type (predictable vs. non-predictable). Overall, the results indicated there was no effect of sentence type at the noun critical region in studies one, two, and three, as well as the verb critical region in studies three and four. However, there was a main effect of sentence type at the noun critical region in study four. Therefore, the hypothesis that reading

times at the critical regions would be faster for the predictable sentences compared to the non-predictable sentences has limited support, with possible explanations discussed below.

The cloze probabilities calculated during sentence norming suggest a difference in predictability across sentence types. Therefore, the absence of a predictability effect is likely not the result of the sentence stimuli. Another possibility is that across age groups, participants were largely not engaging in predictive processing. As described in the introduction, while researchers agree predictive processing during language comprehension exists, there is disagreement over how prevalent it is and at what level (e.g., phonological, grammatical, lexical, etc.) it occurs during naturalistic language processing (Ferreira et al., 2013; Huettig & Guerra, 2019; MacDonald, 2013). Additionally, under certain conditions, engaging in predictive processing might put individuals at a disadvantage if predictions will likely be wrong. The goal of the predictable sentences in the current study was to encourage prediction by constraining the possible nouns following the verb. The goal of the non-predictable sentences was to discourage prediction by using non-constraining verbs, which would increase the likelihood that a participant's predictions would be incorrect. Therefore, since the main effect of sentence type was largely not observed, one must consider why individuals did not demonstrate faster reading times for the predictable condition.

One possibility is that the syntactic construction of the sentences themselves discouraged prediction. In many prior studies that examined prediction, the sentence structure was kept constant across items. For example, in the study conducted by Hintz and colleagues (2017), participants completed a visual world paradigm and heard sentences with the construction "The man _____ at this moment a/an _____.".[1] It is possible that participants adapted to the sentence

[1] This sentence construction is a direct translation of the sentences used in the study, which was conducted in Dutch. While this construction feels unnatural to Native English speakers, it follows Dutch grammar and syntax.

structure in these cases, which as a result, may have decreased task load, freeing cognitive resources for prediction. However, in naturalistic language processing, it is unlikely that individuals encounter the same sentence structure repeatedly. Therefore, in the current project, I aimed to vary the sentence structure to create more ecologically valid sentence items. However, in doing so, this might have discouraged individuals from engaging in predictive processing. This highlights the notion that during natural language processing, individuals likely engage in prediction on a smaller scale because the harm of using cognitive resources to make incorrect predictions might outweigh the faciliatory benefit of correctly predicting during comprehension.

Another possible reason why prediction effects were not observed relates specifically to the study methodology. Prediction studies usually employ eye-tracking or EEG paradigms that can reveal behavioral or processing differences to the stimuli in real time. While other studies have also used self-paced reading paradigms to examine prediction, they are usually in concert with EEG to examine N400 and prefrontal positivity patterns. While there was no behavioral difference in reading times for the predictable and non-predictable sentences in either younger or older adults, it is possible that there were processing differences between conditions that could be observed using EEG. Additionally, in the current study, the sentence stimuli were constructed to encourage lexical-level prediction. As outlined above and in the introduction, researchers hypothesize that during natural language processing individuals likely predict at a more general level (i.e., anticipating a noun vs. a verb). On the other hand, lexical-level prediction requires very specific constraints to occur, because the likelihood of predicting incorrectly is so high. In a visual world paradigm, for example, we often observe lexical-level prediction effects. In a typical visual-world setup, participants are constrained by predictable sentences, as well as by the objects depicted in the display. If participants heard the sentence “The boy peeled the large

____." participants would view four items while hearing the sentence and only one item in the display would be capable of being *peeled.* Therefore, researchers strongly bias participants in favor of prediction because they constrain the possibilities to one. However, using the previous example, in the self-paced reading task, participants were only constrained by the verb. After reading a constraining verb like *peeled,* there were still a handful of possible words that might follow. While participants were constrained to a small number of possibilities in the predictable sentences compared to the possibilities in the non-predictable sentences, the increased likelihood of making an incorrect prediction might have been high enough to discourage lexical-level prediction. Participants may have been predicting at a categorical level (e.g., items capable of being peeled) but the self-paced reading task in the current project was not designed to capture other levels of prediction. Therefore, self-paced reading alone might not be constraining enough to encourage lexical prediction, and using other methods like eye-tracking, which can encourage lexical-level prediction, or EEG to observe neural effects of predictive processing more generally, should be used.

One might also question whether participants were truly engaged with the study since it was conducted entirely online. It is possible that without a researcher present or a distraction-free lab environment, participants may have been less attentive to the task compared to data collected in-person. Additionally, during the self-paced portion of the study, individuals may have simply pressed the button without reading the sentence. However, this scenario is unlikely. While there is less control over the data collection environment in an online study, data quality was assessed, and poor-quality data were discarded. Additionally, the inclusion of yes/no comprehension questions after 20 percent of sentences served to ensure participants were reading the sentences and not just pressing buttons. Both younger (average 90.98% correct across studies) and older

(average 93.66% correct across studies) adults scored well above chance on these questions, demonstrating that they were engaged in the task and reading the sentences.

Self-Paced Reading: Age-Group Differences

Although there were no significant main effects of sentence type, there were age-group differences in reading times at the critical regions across all four studies, supporting the hypothesis that younger adults would have significantly faster reading times than older adults. However, I originally hypothesized that the age-group differences in reading times would reflect younger adults being better at language prediction than the older adults. Considering there were largely no effects of sentence type, nor were there any significant interactions between sentence type and age-group, the faster reading times in younger adults do not indicate more successful prediction. Instead, these differences in reading times may reflect general slowing in the older adults. Further supporting this interpretation is that across the four studies, older adults were significantly slower on the processing speed measures.[2] However, follow-up analyses examining the interaction between processing speed and age-group were not significant.

Alternatively, the main effect of age-group on reading times to the critical region might be explained by participants' strategy during the self-paced task. Participants were instructed that they would be answering comprehension questions following some sentences, but not all sentences. While both younger and older adults performed well on the comprehension questions, the older adults slightly, but not significantly, outperformed the younger adults.[3] While purely speculative, older adults possibly were more attentive to the task and read the sentences more slowly to retain more information to answer the questions accurately. However, without

[2] There was not a significant difference in simple processing speed in study one.
[3] Collapsed across studies, there was no significant difference in comprehension question performance. However, in studies one and two, older adults did answer the questions more accurately than younger adults.

additional information regarding participant strategies during task completion, this interpretation cannot be confirmed.

Language Prediction and Production Relationship

In this dissertation, I aimed to find further evidence supporting the existence of a relationship between language prediction and language production. Examining the results across all four studies, while the findings were varied, the evidence does lend some support to the existence of the relationship. However, the methodology used in the current project, and what the prediction–production relationship signifies, must be considered.

Studies one and two explored the prediction–production relationship across age groups using offline measures of language production to predict reading times during the self-paced tasks. Therefore, these studies were assessing if participants' general language production abilities were related to their language prediction performance. On the other hand, studies three and four included online language production assessments to examine if the expectation to produce would encourage engagement in language prediction. Taken together, the results do not support a relationship between general language production abilities influencing language prediction—with a few caveats to this interpretation discussed below—however, when participants were aware they must produce, the evidence suggests greater engagement in language prediction.

Across all four studies, the language production composite score did not predict reading times at the critical region. At first glance, one might interpret this as evidence against the existence of the prediction–production relationship, or at the very least, that the relationship is weak. However, as described in the introduction, prior studies have found evidence in favor of the relationship. For example, toddlers with larger productive vocabularies had more looks to the

target item in a preferential looking paradigm (Mani & Huettig, 2012) and younger adults with higher verbal fluency showed greater prediction effects during a visual world paradigm (Hintz et al., 2017). Perhaps most compelling, during an EEG study, when participants' production system was occupied, there was no N400 effect, suggesting that the production system is required to engage in prediction (Martin et al., 2018). This leads to the question, if prior research found evidence in support of the relationship between language prediction and language production, why was there no main effect of language production composite score during the self-paced reading task?

One possibility for not finding a main effect of language production composite score on prediction is that participants were not engaging in prediction during the self-paced tasks. If prediction did not occur, one would not expect to find a relationship between language production composite scores and reading times. While this may be true for studies one through three, there was a main effect of sentence type in study four, suggesting that participants did engage in language prediction. Alternatively, the language production composite score may have captured participants' general language production abilities, but language prediction might be more strongly related to certain aspects of language production. For example, researchers in the studies mentioned above found a relationship between language prediction and verbal fluency, which Hintz and colleagues (2017) termed production fluency (Federmeier, et al., 2010; Hintz et al., 2017). Therefore, specific production measures, like verbal fluency, might be more strongly linked to language prediction.

This interpretation is further supported when examining the results within the production blocks of studies three and four. The language production composite score was not predictive of picture naming accuracy and picture naming response times, both of which were measures

included in the language production composite score. Therefore, since the language production composite score did not predict language production performance, using a composite score that includes varied metrics of language production is likely not the best approach to examine the prediction–production relationship. Using the composite score was a strategy employed to prevent issues of multiple comparisons; however, collapsing across metrics likely removed some of the variability and nuance we often observe on individual measures of language production. Consequently, in future studies, instead of using a single composite score to evaluate the prediction–production relationship, focus should be placed on metrics examining production fluency, which might include verbal fluency, as well as participants' number of disfluencies (e.g., number of repetitions, pauses, elongations, filler words, etc.) during naturalistic production, which would require overt speech articulation as opposed to typed production.

Although the results examining the effect of language production composite score on language prediction score were not significant, I also examined whether giving participants the expectation to produce would encourage predictive processing. Results from the two production blocks provide support for the relationship between language prediction and production, with more consistent results observed when examining picture naming response times. In both studies, the predictable items were named faster than the non-predictable items. Additionally, in study four, predictable items were named more accurately than non-predictable items. Individuals responding faster—and more accurately in study four—to predicted items compared to non-predicted items suggests that individuals were generating expectations and engaging in language prediction. This finding supports the hypothesis that providing participants with the expectation to produce will encourage language prediction engagement.

Considering the results from the production blocks, results across studies one and two, as well as results from the comprehension block in study three, one explanation for the disparity in prediction results across studies is that during the comprehension-only tasks, participants were less inclined to engage in prediction because the benefits of predicting were small. While prediction is not required for comprehension, it may facilitate it. However, the cognitive resources required to predict are costly. Therefore, in a situation when the sentences are relatively simple and straightforward, like in the current study, facilitating comprehension may be unnecessary, with potential comprehension benefits from predicting not worth the cost of using the extra cognitive resources. On the other hand, if participants were engaging in language production, then their production system was already active and may have primed participants to predict. More specifically, the production system, which is hypothesized to be the same system used in prediction, was already in use for the production block task. One hypothesis is that the cognitive burden of activating the production system during the picture naming portion, and then deactivating it during the comprehension portion of each trial is high. To ease cognitive load, participants may have had their production system active during the entirety of the production block and utilized it to make predictions during the comprehension portion of each trial. Therefore, participants were incentivized to predict because it would improve their task performance, and the costs of predicting were lower because the production system was already engaged for the naming task.

If participants were in a predictive state due to the activation of their production system for the duration of the block, this would explain the differences in naming response times between sentence types. In study three, there were no accuracy differences in naming across sentence types, suggesting that participants were error monitoring and correcting incorrect

predictions before responding. Because the predictions made during the predictable sentences were more likely to be confirmed by the picture, fewer prediction errors would occur, resulting in faster response times. On the other hand, predictions made during the non-predictable sentences were more likely to be disconfirmed by the picture, which suggests that to maintain high response accuracy, participants were monitoring their prediction errors, resulting in slower response times for the non-predictable items. I will explore the ramifications of this across age groups in the next section.

In addition to encouraging prediction through production engagement in the production blocks, studies three and four also aimed to determine if there were differences in prediction engagement by comparing prediction performance across blocks. Reading times at the verb were used for these analyses because the verbs were consistent across blocks. Results across studies three and four indicated that participants' reading times were faster at the verb critical region in block two compared to block one. This might suggest that production engagement during block two increased prediction more generally, and that participants were not only predicting the target items. Because results from the comprehension block in study three do not suggest participants were engaged in prediction, I focused on study four.

Results from study four revealed a main effect of sentence type on the noun critical region, with the predictable critical region read faster than non-predictable critical region. There was also a main effect of sentence type on picture naming accuracy and response times in the production block. Therefore, to compare prediction across blocks, I examined the verb critical region. Previously, I mentioned that reading times at the verb critical region were significantly faster in the production block compared to the comprehension block, which could suggest more generalized prediction in block two as a consequence of production engagement. However,

across participants, the results do not suggest engagement in generalized prediction but instead suggest practice effects.[4] To arrive at this conclusion, I conducted follow-up tests within the language production block. If it were true that engagement in language production encouraged language prediction more generally, then performance on the picture naming task should be associated with reading times at the verb critical region. Although naming response times and reading times at the verb critical region had a significant positive correlation, this likely reflects processing speed more generally. I also tested whether picture naming response times predicted reading times under the assumption that if production encourages prediction, then production performance should predict the reading times. However, this was not the case. Finally, as previously mentioned, the main effect of block was found in the absence of a main effect of prediction or an interaction between block and sentence type. This suggests that the faster reading times at the verb critical region were not associated with predictive processing and instead reflected order or practice effects. Because participants always completed the comprehension block first, I could not test for order effects across blocks. However, because trial order was pseudorandomized to ensure the comprehension questions occurred with the correct sentence, all participants within each counterbalance list completed trials in the same order. I conducted a paired t-test across participants comparing reading times at the critical region to the verb for the first half of trials in the production block to the second half of trials. I found that reading times for trials that occurred during the second half of the block were significantly faster than reading times for trials during the first half of the block. Therefore, the effect of block does

[4] A caveat to this conclusion is that there might be some individual differences or age group effects in regard to generalized prediction, which are discussed in the section below specific to age group differences in language prediction and production. However, this does not negate finding evidence of practice effects.

not suggest that the production block encouraged more general engagement in prediction but reflects practice effects, with participants responding faster over the duration of the experiment.

This project did not find conclusive evidence supporting production engagement encouraging general predictive processing over and above circumstances when the language production system was not engaged. However, the evidence from the production block does support that the relationship between language prediction and language production exists. Not finding strong evidence for generalized prediction at the verb across participants is not necessarily surprising when considering the nature of the production task coupled with the evidence in support of prediction at the target noun. To expand on this, participants were engaged in a goal-directed task and knew they would be naming pictures and were instructed to do so quickly and accurately. Therefore, when they were predicting, efforts were likely targeted toward anticipating what the upcoming picture might be to improve task performance. One can think of this in terms of the results found in the study conducted by Martin et al., (2018). In this study, when the production system was occupied, participants were unable to engage in predictive processing. Relating these results to the current study, it is possible that during the production block, participants' production system was engaged in service of the task. As discussed above in the section on language prediction, the comprehension task alone likely was not constraining enough to encourage lexical-level prediction. However, participants knew they would have to identify a single word (or two) to name the picture item. Therefore, providing participants with the expectation to produce encouraged lexical-level prediction. Hence, participants' production system was occupied as they attempted to anticipate the upcoming picture item to name it as quickly and accurately as possible. Consequently, with the task demands, the prediction system (i.e., the production system) was not available to participants to

generate predictions more broadly about any possible upcoming information. The current results, therefore, do not find evidence in favor of language production engagement encouraging language prediction more broadly, but the possibility cannot be entirely discarded, either. A different type of language production manipulation may encourage broader, or more general, language prediction, if the task warrants it. However, even if researchers can use production to encourage more general prediction, it is unlikely individuals will anticipate all upcoming information because this would be maladaptive and impede comprehension.

Prediction is a tool individuals can engage during comprehension when circumstances favor it. Therefore, engagement in production does not necessarily mean participants will actively predict all upcoming information, especially because a large proportion of those predictions would likely be incorrect. Instead, having participants engaging their production system might have them "prediction ready," and would make engagement in prediction less effortful if circumstances are favorable. Interestingly, when looking at individual naming responses across participants, I observed that certain items were named incorrectly more often when preceded by the non-predictable verb than the predictable verb, suggesting failed error monitoring and updating. While analyses looking at specific responses were outside the scope of this dissertation, I noticed during data coding that while certain items, like *banana* or *bike* were usually named correctly regardless of condition, some items showed more naming variability. For example, when preceded by the predictable verb *mine (for)*, participants typically responded correctly with *gold*. However, when preceded by the non-predictable verb *look (for)*, the responses were more diverse, with some participants responding correctly, while others responded with *stone* or *rock*. Other items showed a similar pattern of responses. This might suggest that for predictable sentences, participants were generating expectations that were

confirmed by the picture, whereas the non-predictable sentences were either not constraining enough to encourage participants to generate expectations, or if participants were predicting, did not always effectively update their disconfirmed predictions correctly (or as specifically as required—i.e., responding *rock* for the picture of a gold nugget). Additionally, this observation is consistent with the accuracy results in study four in that the predictable items were named more accurately than the non-predictable items. Prior research suggests that individuals are more likely to engage in prediction with more non-linguistic scaffolding. I manipulated this by increasing the amount of semantic context participants were provided. Based on the results from the production blocks described thus far, engagement in production might serve a similar function to non-linguistic scaffolding by providing individuals with support to engage more readily in prediction.

This series of studies investigated language prediction and language production. Overall, the results—especially those from studies three and four—support a relationship between language prediction and production. However, current results suggest that this relationship is not necessarily a causal relationship, or a perfect 1:1 relationship. Instead, encouraging individuals to engage in language production may serve as a form of non-linguistic support. If language prediction uses the same system as language production, and the production system is active, participants might be more likely to predict. In addition to exploring the relationship between language prediction and language production, the other main question of interest was whether there are age-related differences in the relationship between language prediction and production.

Age-Related Differences in Language Prediction and Production

Because the aging literature finds evidence for declines in language production, as well as in language prediction, we must ask if declines in these areas influence the relationship between language prediction and language production in older adults. In the current project, the results

revealed some general age-related differences in performance. This includes the differences in general language production performance, as well as differences in reading times, both of which are reported above. However, the question remains whether these age-related declines are associated with differences in the language prediction and production relationship between younger and older adults. I hypothesized that declines in these areas would translate to age related differences in the prediction–production relationship. While some evidence supports this hypothesis, the differential effects of language prediction and language production were not consistently observed. This suggests that the strength of the relationship might vary across older adults and is possibly related to their overall cognitive functioning.

In exploring the existence of age-related differences in the relationship between language prediction and production, I turn again to studies three and four. During the production block in study three, there was a significant age-group by sentence type interaction. While both age groups showed significantly faster picture naming response times for predictable items compared to non-predictable items, the older adults showed greater performance improvements when naming predictable items compared to the younger adult group. This suggests that the relationship between language prediction and language production exists in both younger and older adults. However, it also suggests that while older adults can benefit from predicting, making incorrect predictions is more costly.

As discussed above, to maintain similar response accuracy across sentence types in study three, participants would have to monitor their predictions and make updates when these predictions were wrong. For younger adults, they likely made these updates more quickly when their predictions for non-predictable items were incorrect due to a higher degree of available cognitive resources. On the other hand, due to age-related declines in cognition, older adults

would likely require more time to monitor their prediction errors and make updates. Additionally, older adults have larger semantic knowledge stores than younger adults, so their predictions during the non-predictable sentences may have been more diverse than younger adults (Park et al., 2002). Potentially, their incorrect predictions for non-predictable items may have been a larger distance away from the correct item in their semantic network, necessitating more time to access the correct item name than younger adults needed. Consequently, the higher cognitive demands older adults may experience to monitor and update their prediction errors may lead to their decline in prediction engagement more generally because the cost of making incorrect predictions is higher than any benefits gained from predicting.

While previous research suggests that older adults are less likely to engage in prediction, they may be encouraged to do so through non-linguistic scaffolding. While the younger adults were already likely performing close to ceiling across sentence conditions, the older adults' engagement in language production may have served as a form of non-linguistic scaffolding. Therefore, for the older adults, encouraging them to engage their production system might have led them to predict, because it would benefit their performance in a goal-directed task.

I also looked across blocks to determine if there were age-related differences in the prediction–production relationship. I examined the interaction between sentence type and age group, as well as their main effects, on the prediction score calculated at the verb. While differences in reading times across blocks can likely be attributed to practice effects, the prediction score examined the differences in reading times for each participant between the non-predictable and predictable conditions.

Before exploring the relationship across blocks using the verb critical region, as a reminder, in study four, there was a main effect of sentence type on reading times at the noun

critical region in the comprehension block, as well as on the picture naming response times in the production block. Furthermore, the unexpected finding during the moderation analyses indicate that older adults actually had higher prediction scores at the noun critical region, suggesting the older adults may have driven the main effect of prediction during block one. Taken together, this indicated that the older adults were engaging in language prediction across study four.

Returning to the prediction scores at the verb to examine prediction engagement across blocks, although the older adults were predicting across study four, there was a main effect of age group on prediction scores calculated at the verb, with younger adults having higher prediction scores. This might suggest that younger adults were engaged more in prediction during the production block than the older adults. To investigate this further, I conducted the same analysis at the verb critical region within each block. There was no effect of age group on the prediction scores in block one, but an effect of age group was found in block two. However, the average prediction scores within each age group did not significantly differ across blocks.

Examining the datapoints visually suggests that at the individual level, the differences between reading times for the non-predictable and predictable sentences appears to have decreased in younger adults, with overall variability in prediction score declining. On the other hand, in the older adults, it appears that more individuals had negative prediction scores, and the overall variability in prediction scores appears to have increased. I can only speculate, but maybe the added context in study four increased the predictability of the non-predictable items, resulting in more negative prediction scores within the older adults. Alternatively, while the presence of a significant age group effect is apparent, the reading time differences resulting in the prediction score were calculated at the millisecond level. Therefore, it is questionable whether an average

reading time difference of 6 ms or 20 ms will have a meaningful effect on behavior. However, the possibility that individual differences might exist cannot be discounted.

While the results seem inconsistent, they do highlight that the relationship between language prediction and production might be influenced more by cognitive factors like working memory capacity, processing speed, or reading experience, than by age. Although older adults often show declines in these areas, across all four studies there were no significant performance differences on the working memory tasks between younger and older adults, except in study four, in which older adults scored significantly higher on the forward digit span task. Additionally, the older adults outperformed younger adults on both reading experience measures (author recognition and vocabulary). Therefore, while evidence from study three does suggest that the relationship between language prediction and production might differ between younger and older adults, as evidenced by a significant interaction between age-group and sentence type, the older adults were still engaged in prediction. To disentangle this relationship, further research needs to be conducted and include older adults who demonstrate more variability in cognitive performance.

Self-Paced Reading: Effects of Context

Establishing the existence of the relationship between language prediction and language production, and whether this relationship differed with age, were the main aims of the current project. However, the effect of non-linguistic scaffolding was also explored. Between-subjects analyses were conducted to determine if the added context in studies two and four improved performance compared to studies one and three, respectively. The results of both analyses were not significant. Furthermore, the hypothesis that there would be an interaction between sentence type and context on reading times was not supported. This hypothesis assumed participants

would be engaging in predictive processing, indicated by significant differences in reading times across sentence types. However, this was not the case across studies one and two, in which there was no main effect of sentence type in either study. While there was a main effect of sentence type in the comprehension block for the noun critical region in study four, the main effect of context was not significant. There was also no interaction between sentence type and context across studies three and four. Since the results from the comprehension block of study three do not indicate that prediction occurred, and the difference in reading times between conditions in study four was small, though significant, not finding the interaction was unsurprising.

Additionally, the hypothesis that there would be a sentence type by age-group interaction was not supported. Although older adults are less likely to engage in language prediction compared to younger adults (Federmeier & Kutas, 2005; Federmeier et al., 2010), they might be encouraged to predict with added context. Therefore, I hypothesized that younger adults would already be performing close to ceiling and that older adults would benefit more greatly from the added context. However, as already stated, the evidence that predictive processing occurred during the comprehension-only portions was weak, and there is no evidence in the current studies that older adults benefitted more from context than the younger adults. Although these findings suggest that the added context did not significantly benefit performance compared to the no-context condition, when examining the results within studies two and four, there is some evidence that prediction took place. First and foremost, within study four, there was a significant effect of sentence type at the noun critical region, with faster reading times for the predictable critical region compared to the non-predictable critical region. This effect was not observed in studies one through three and may indicate that overall, participants were encouraged to predict

because of the added context. This leads to the question of why this same effect was not observed in study two.

When examining the effect of prediction in study two, several moderation analyses were conducted to determine if the relationship between language prediction and production was moderated by cognitive factors such as working memory, processing speed, or reading experience. While the results and ramifications of these analyses are discussed in the section below, one finding was that there was a main effect of working memory and reading experience on prediction score irrespective of age, with higher scores on both cognitive factors associated with higher prediction scores. As a reminder, prediction scores were calculated for each participant by subtracting the average reading time for predictable sentences from the average reading time for non-predictable sentences. Therefore, larger, positive values suggest an effect of prediction. Although there was no main effect of sentence type on reading times in the main analysis, these results suggest that participants with higher working memory capacities and more reading experience were engaging in prediction, possibly due to the added semantic context. This is consistent with prior studies that found working memory (Federmeier et al., 2005; Huettig & Janse, 2016) and reading experience (Ferreira, Foucart, & Engelhardt, 2013; Huettig & Pickering, 2019) as underlying mechanisms of language prediction. One interpretation is that as context was added, participants were required to keep more information online to answer any potential comprehension questions following the sentences. Therefore, individuals with lower working memory scores and less reading experience may have reached their working memory capacity and did not have enough cognitive resources available to engage in prediction. Individuals with higher working memory capacity and who read more, which has the benefit of improving working memory and training the prediction system, had enough resources available

to utilize the semantic context to predict and facilitate their comprehension of the sentences, possibly in service of answering the comprehension questions more accurately.[5]

Underlying Mechanisms

As discussed above, language prediction is not a requirement of language comprehension. Individuals therefore might only engage in predictive processing when the circumstances favor the accuracy of their predictions. Additionally, when and if individuals engage in language prediction may vary as a function of their cognitive resources. Previous studies have found a relationship between working memory and processing speed, irrespective of age, with higher working memory and processing speed abilities associated with engagement in prediction (Federmeier et al., 2005; Huettig & Janse, 2016; Huettig & Pickering, 2019). Furthermore, because reading engagement can train the prediction system, as well as help improve working memory abilities, more reading experience is associated with more predictive processing engagement.

Cognitive factors such as working memory and processing speed also have a relationship with language production. As such, there is a possibility that the strength of the relationship between language prediction and production varies across individuals according to cognitive ability, with individuals who score higher on cognitive tasks more likely to engage in prediction when language production is expected. To investigate the possible relationship cognitive factors have with language prediction and production, moderation analyses were conducted to determine

[5] This pattern of results on prediction score was not observed in study four. One possibility was that the individuals in study four had higher working memory capacity and more reading experience than participants in study two. Independent samples t-tests, however, do not support this interpretation. The likely explanation is that while the moderation analyses that detected main effects of working memory and reading experience were conducted irrespective of age, in study four, there was a main effect of age group on prediction score, which is discussed in the section examining the relationship between language prediction and production.

if varying levels of working memory, processing speed, and reading experience moderated the relationship between language prediction and language production across age groups.

Overall, across all four studies, there was no evidence supporting the hypotheses that the relationship between language prediction and production is moderated by working memory, processing speed, or reading experience. However, an absence of supporting evidence in the current project does not necessarily mean the relationship does not exist. As discussed above, the possibility remains that in studies one and two and during the comprehension block for study three, participants did not engage in prediction. Additionally, in study four, where there was a main effect of sentence type suggesting engagement in prediction, the mean difference in reading times across sentence types differed by less than 25 ms. Therefore, when examining individual differences for the moderation analyses, there may not have been enough variation to detect the relationship. Finally, as previously discussed, the language production composite score used in the project may not have been the best approach because it did not predict performance in any analyses. The overall composite score may not have captured the aspects of production related to prediction, and using a different measure of language production, like verbal fluency, might serve as a better predictor in the model. Taken together, this further supports the notion that using more robust methods during prediction studies might be necessary to understand underlying cognitive mechanisms of the relationship between language prediction and production.

Future Directions

Although efforts were made during the study design and data collection process to encourage language prediction and evaluate the relationship between language prediction and production, several aspects of the project can be improved upon in future studies to increase the

likelihood of observing predictive processing, its relationship with production, as well as underlying mechanisms. Some areas to improve include changes to the experimental methods, which language production tasks to use, and using a within-subjects design. Additionally, implications of prediction studies on natural language processing are also considered.

Although the sentence norming conducted for this project indicated significant differences in cloze probabilities between the predictable and non-predictable sentences, prediction effects were not always observed. As discussed in the section above on language prediction performance, it is possible that the self-paced reading task was not constraining enough to encourage lexical-level prediction. Additionally, it was not designed to detect prediction that may have occurred at other levels. Moreover, self-paced reading cannot detect differences in neural processing that may be occurring during sentence reading, even in the absence of behavioral differences. Therefore, future studies examining the relationship between language prediction and language production should employ EEG methods.

In addition to the study method, the language production tasks used should also be re-examined. In the current studies, several different structured and naturalistic tasks assessing language production were used. Results on these measures were then combined into one score, which was used in all subsequent analyses. However, it is possible that certain aspects of language production are more highly related to language prediction than others, especially when considering that language prediction is not a requirement of language comprehension. For example, if participants are more likely to engage in prediction in the presence of additional non-linguistic scaffolding, such as more time or a richer semantic context, then more naturalistic measures of language production that measure production disfluencies and require participants to produce more complex syntactic utterances should be used in future studies.

In addition to changes in the methods, future studies should incorporate a within-subjects design to determine if participants show improvement in prediction performance with added sentential context. While the current project looked at differences in performance across contextual conditions, results did not suggest greater predictive processing engagement with added context. However, these analyses were conducted between-subjects, and the benefits of contextual support might be observable within-subjects, with individuals showing varying levels of engagement in prediction depending on the presence of a richer semantic context.

I repeatedly highlight throughout the discussion that language prediction is not necessary for comprehension but may facilitate it under beneficial circumstances. Additionally, results in the current project indicated that engagement in language production can encourage lexical-level prediction. However, one question to ask in future studies is what does this imply for naturalistic language processing? Previous studies already suggest that people likely rarely engage in lexical-level prediction because the risk of incorrect predictions is high unless conditions are highly constrained. However, many people can think of an occasion when they predicted upcoming information, whether it be in a conversation or during a movie. Therefore, to understand when and how people predict during natural language contexts, researchers might consider conducting discourse studies, in which predictability is manipulated, and asking more qualitative questions about participants' own awareness of their predictions.

Conclusion

In this dissertation, I aimed to answer two overarching research questions: (1) is there a relationship between language prediction and language production and (2) if this relationship exists, does it differ across age groups. In a series of four self-paced reading studies, I found evidence supporting the existence of a relationship between language prediction and language

production, as well as some evidence in support of there being age-related differences in this relationship, with the caveat that finding differences may depend on individual cognitive abilities of participants.

In addition to exploring the relationship between language prediction and language production, I was also interested in exploring potential underlying mechanisms of the relationship. I assessed whether cognitive factors including working memory, processing speed, and reading experience moderate the prediction–production relationship. However, I did not find any significant moderation effects, possibly due to a lack of individual variability across these metrics, or that in studies one and two, I did not find evidence that prediction occurred.

I also examined whether differences in the amount of non-linguistic scaffolding impact prediction. Overall, results did not support that a richer semantic context encouraged prediction over and above the studies that did not have additional contextual support. However, within study two, results suggested that higher working memory abilities and more reading experience were associated with higher prediction scores, suggesting that with the added context, individuals with greater cognitive resources at their disposal were encouraged to predict.

Lastly, results suggest that having participants engage their language production system encourages prediction. One possibility is that activating the production system serves a similar function to other types of non-linguistic scaffolding. When expecting to produce for a goal-directed task, participants might be encouraged to predict to improve task performance. Additionally, if participants expect to produce, they might keep their language production system engaged throughout the task, as repeatedly engaging it to produce, and disengaging it during comprehension would be cognitively demanding. If the production system is already activated during comprehension, participants might be more likely to engage in prediction because it may

benefit their task performance. However, further research specifically examining neural correlates of language prediction and production is needed to confirm this hypothesis. Furthermore, in-person research using overt speech measures of language production to examine more nuanced aspects of language production, such as production fluency, are needed to determine whether general language production abilities are related to language prediction, or if some components of production are more strongly associated with prediction engagement.

While the current project cannot conclusively state that prediction *is* production, the evidence is consistent with prior studies suggesting a relationship between both processes, and that there may be some age-related differences in the relationship, possibly due to individual variability in cognitive abilities.