Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

2024, Vol. 50, No. 1, 109–136
https://doi.org/10.1037/xlm0001273

The Role of Cognitive Control and Referential Complexity on Adults' Choice of Referring Expressions: Testing and Expanding the Referential Complexity Scale

Madeleine Long[1, 2], Hannah Rohde[3], Michelle Oraa Ali[4], and Paula Rubio-Fernandez[1]
[1] Department of Philosophy, University of Oslo
[2] Multimodal Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[3] Department of Linguistics, School of Philosophy, Psychology and Language Sciences, University of Edinburgh
[4] Department of Cognitive, Linguistic and Psychological Sciences, Brown University

This study aims to advance our understanding of the nature and source(s) of individual differences in pragmatic language behavior over the adult lifespan. Across four story continuation experiments, we probed adults' ($N = 496$ participants, ages 18–82) choice of referential forms (i.e., names vs. pronouns to refer to the main character). Our manipulations were based on Fossard et al.'s (2018) scale of referential complexity which varies according to the visual properties of the scene: low complexity (one character), intermediate complexity (two characters of different genders), and high complexity (two characters of the same gender). Since pronouns signal topic continuity (i.e., that the discourse will continue to be about the same referent), the use of pronouns is expected to decrease as referential complexity increases. The choice of names versus pronouns, therefore, provides insight into participants' perception of the topicality of a referent, and whether that varies by age and cognitive capacity. In Experiment 1, we used the scale to test the association between referential choice, aging, and cognition, identifying a link between older adults' switching skills and optimal referential choice. In Experiments 2–4, we tested novel manipulations that could impact the scale and found both the TIMING of a competitor referent's presence and EMPHASIS placed on competitors modulated referential choice, leading us to refine the scale for future use. Collectively, Experiments 1–4 highlight what type of contextual information is prioritized at different ages, revealing older adults' preserved sensitivity to (visual) scene complexity but reduced sensitivity to linguistic prominence cues, compared to younger adults.

Keywords: language production, reference, aging, attention switching

Supplemental materials: https://doi.org/10.1037/xlm0001273.supp

Over the last half-century, interest has grown in understanding the extent to which aging affects linguistic skills, given the strong association between social interactions and elderly life satisfaction (Lubben & Gironda, 2003). While it is commonly assumed that the domain of language is well-preserved across the adult lifespan, emerging research suggests that not all aspects of language are equally spared with advanced age (Wright, 2016). One area prone to decline is *pragmatics* (i.e., the socio-cognitive skills that allow us to infer meaning beyond what is said and tailor our utterances to our interlocutor; Messer, 2015).

The cause of the pragmatic decline in healthy aging is unclear. A number of pragmatic skills (e.g., efficient communication, perspective-taking, and the use/interpretation of figurative language) have been linked to *executive functions* (i.e., the set of cognitive skills associated with goal-oriented actions; Antoniou et al., 2016; Bambini et al., 2021; Long, Rohde, & Rubio-Fernández, 2020;

This article was published Online First August 10, 2023.
Madeleine Long https://orcid.org/0000-0002-6034-4038
Michelle Oraa Ali https://orcid.org/0000-0002-3653-8078
The authors would like to thank Merel Scholman for providing the automated reading span task along with the relevant R script to extract participant scores. The research reported in this study was supported by the Young Research Talent Grant from the Research Council of Norway awarded to Paula Rubio-Fernández (Saying enough without saying too much—Project 230718) and a Leverhulme Trust Prize in Languages and Literatures to Hannah Rohde. The authors gratefully acknowledge this funding.

All materials, data, and analyses have been made publicly available on OSF and can be accessed at: https://osf.io/7dzta/?view_only=04a2c177e2564b27a6699cde4ce5bbdd.

The data are available at https://osf.io/7dzta/?view_only=04a2c177e2564b27a6699cde4ce5bbdd

The experimental materials are available at https://osf.io/7dzta/?view_only=04a2c177e2564b27a6699cde4ce5bbdd

Correspondence concerning this article should be addressed to Madeleine Long, Department of Philosophy, University of Oslo, Blindernveien 31, Georg Morgenstiernes hus, 0313 Oslo, Norway. Email: madeleine.long@ifikk.uio.no

Long et al., 2018). These cognitive skills follow a similar lifespan trajectory as pragmatic skills, developing in childhood, peaking in young adulthood, and deteriorating in old age (e.g., Messer, 2015), which has led researchers to theorize that executive functions underlie the processing and production of pragmatic-relevant information. However, this is not the case for all pragmatic skills (e.g., Schubotz et al., 2019), which suggests that certain skills may depend less on domain-general cognitive control and more on language-specific mechanisms and/or social and environmental factors. Collectively, these findings highlight the complex relationship between language, cognition, and aging and the need for further investigation.

From both a theoretical and practical standpoint, the investigation of this complex relationship has the potential to push the field forward. On a theoretical level, we can gain a unique insight into the human mind by studying older adults as they have extensive language experience but vary considerably when it comes to cognitive control. Unlike children (whose linguistic and cognitive skills are still developing) and young adults (whose linguistic and cognitive skills are typically at ceiling), older adults' distinct profile offers an ideal testbed for examining the extent to which the ability to regulate language reflects cognitive control or sensitivity to other factors in the discourse context. By conducting comprehensive studies over the adult lifespan, we can thus better understand the dynamics between language skills and cognition at different stages of development.

On a practical level, identifying linguistic patterns in healthy aging is a prerequisite to identifying linguistic patterns in pathological aging (Wright, 2016). By pinpointing the conditions under which younger and older adults' behavior diverges, we can provide the necessary normative data to inform clinical practice regarding communicative health and interventions and to promote successful social interactions in old age, with the ultimate goal of empowering elderly individuals and improving life satisfaction.

Against this background, the current study aims to advance our understanding of pragmatic language behavior in aging through the contribution and analysis of a large set of data on adults' referential choices over the lifespan ($N = 496$ participants, ages 18–82). Here we systematically probed younger, middle-aged, and older adults' choice of more or less explicit referential forms (i.e., proper names vs. pronouns to refer to the main character) across four story continuation experiments. As pronouns are a discourse marker[1] of topic continuity (i.e., a way of marking that the discourse will continue to be about the same referent; Sandoz et al., 2023), the choice of whether to use a pronoun or proper name provides insight into participants' perception of the topicality[2] of a referent in a given context.

Our contextual manipulations were based on Fossard et al.'s (2018) scale of referential complexity, whereby the complexity of coreference decisions is classified according to the visual properties of the scene: specifically, the number of characters (one or two) and their sex/gender (same or different). In Experiment 1, we used Fossard et al.'s scale to test the association between referential choice, aging, and cognition. In Experiments 2–4, we expanded on this by testing novel manipulations that could impact referential complexity due to COMPETITION FOR TOPICHOOD. Specifically, we manipulated: (a) the NUMBER of competitor referents in the scene/discourse (0–2); (b) the TIMING of competitors' presence (early or late in the scene/discourse); and (c) the EMPHASIS on competitors (relative to the main character). Overall, our results shed light on the nature and source of age-related differences in referential choice and pinpoint novel factors relevant to referential complexity, which we situate onto Fossard et al.'s complexity scale for the first time, creating a more nuanced continuum.

Referential Choice in Aging

As mentioned above, speakers make referential choices when deciding on the type of linguistic form to use to refer to a given entity. Speakers can choose from a variety of forms to refer to the same entity, from semantically rich proper names, to less explicit pronouns, or physical descriptions. For example, a speaker who introduces a referent "Sally" during a conversation might choose, in subsequent sentences, to refer back to Sally using a reduced form "she." This type of anaphoric reference would work well in contexts where Sally is the only referent being discussed or where the other referents are males. However, when there are multiple female referents being discussed, the use of "she" could create ambiguity for the listener, and the more informative referential form "Sally" may be more appropriate.

A large body of work has investigated speakers' choice of referring expressions in contexts like the one above, where referential expressions can either facilitate the sharing of information with a communicative partner or leave them guessing (e.g., Arnold & Griffin, 2007; Fukumura et al., 2010, 2013; Novogrodsky & Edelson, 2016; Serratrice, 2013). One of the most prevalent findings from this work is that children and older adults have difficulty with anaphora resolution and are more likely to produce ambiguous pronouns than younger adults (Hendriks & Spenader, 2006; Hendriks et al., 2008; Karmiloff-Smith, 1985; Light & Capps, 1986; Light et al., 1994; Pratt & MacKenzie-Keating, 1985; Pratt et al., 1989). In a novel line of research, Hendriks et al. (2014) set out to address what underlies these production differences.

In their study, Hendriks et al. found that children and older adults' use of ambiguous pronouns is related to the complex interplay between topic shifts and discourse stages, which involves making calculations with regard to what information is in the common ground and which referents occupy a status of prominence at any given time. They uncovered this pattern by presenting children, young adults, and older adults with a series of panels in which two characters of the same sex/gender carried out various actions, and asked participants to recount a story for an addressee who could not see the panels. The panels depicted a sequence of events which involved the use of three discourse stages: Introduction, Maintenance, and Re-Introduction. As predicted, calculating what information is in the common ground, particularly when shifting topics, can be mentally taxing for children and older adults whose cognitive resources are not fully developed or could be in decline. What remains unclear, however, is whether such differences also emerge in the absence of topic shifts, when the complexity of the immediate visual context varies systematically.

[1] Note that the above definition of "discourse marker" is one of several definitions used in the literature (see, e.g., Renkema & Schubert, 2018; Shukla et al., 2022 for other uses of the term).

[2] Here and throughout the paper we use the term "topic" to mean the referent who the discourse is about (using the information structural sense of topical "aboutness"; e.g., Lambrecht, 1994).

Aging, Referential Complexity, and Discourse Stage

A promising new line of research has begun to address this question by testing referential choices across all three discourse stages (Introduction, Maintenance, and Re-Introduction) and all three levels of referential complexity (Fossard et al., 2018; Sandoz et al., 2020, 2023). According to Fossard et al. (2018, p. 5), "Referential complexity can increase either when there is an increased number of characters or when two (or more) characters have the same gender." More specifically, Fossard et al. classify these manipulations along the following scale: low complexity level (one character in the scene), intermediate complexity level (two characters of different genders in the scene), and high complexity level (two characters of the same gender in the scene, creating potential ambiguity).

In their study, Fossard et al. (2018) tested younger adults in a storytelling-in-sequence task and found that the referential complexity of the scene affected pronominal use not only at the topic shift (or Re-Introduction stage), which has been the focus of much of the literature, but also at the Maintenance stage where pronominal rates decreased in response to an increase in the number of characters (or referential competitors) in the visual scene. This finding highlights the importance of expanding our focus to other discourse stages which could similarly be affected by referential complexity, and which might interact with age.

Building on this work, Sandoz et al. (2020) replicated findings from Fossard et al. (2018) in Alzheimer's patients and older adult controls and discovered that cognitive resources are not only relevant for Re-Introduction (where topic shifts occur), but also other discourse stages where it may be optimal to signal the status of a referent as topical through reduced referential forms. Specifically, for both older adult and clinical populations, Theory of Mind and "planning skills" (i.e., a set of skills used to carry out goal-directed behavior[3]) were negatively associated with the use of definite expressions in Maintenance contexts (i.e., poorer cognitive skills were associated with the use of a nonoptimal overexplicit referential form for a discourse context involving topic continuity). In interpreting these results, Sandoz et al. expand on the link between planning skills and optimal referential choice by suggesting that "planning abilities could reflect the participant's capacity to guide their addressee to maintain his focus on the intended referent" (p. 16). Since pronouns are reserved for highly topical referents (e.g., Ariel, 1990), the authors argue that the use of other referential forms could disrupt the listener's representation of the active referent, inadvertently signaling a topic shift. As such, those with better planning skills may rely on those cognitive abilities to guide their choice of appropriate referential form.

Expanding on this work, Sandoz et al. (2023) again found that planning skills were predictive of pronominal use in Maintenance contexts (such that better planning was associated with an increase in pronominal use), this time in a group of older adults varying in age from 60 to 91 years old. This finding lends further support to their claim that planning abilities reflect participants' capacity to mark topic continuity for the listener through pronominal use. As few studies have investigated the topic-marking function of pronouns, this work highlights the need for more research on pronouns as a signal of referent status and an indicator of discourse planning, which will be addressed in the current study.

From an aging perspective, the link found between pronoun use and cognitive control was also particularly noteworthy. Given that planning skills predicted pronominal use in the elderly population, it is reasonable to ask whether the same is true for younger and middle-aged adults, whose cognitive functions vary considerably from older adults. Indeed, recent work has shown that depending on the pragmatic ability, younger and older adults rely on the same or different cognitive mechanisms to guide referential communication (Long, Rohde, & Rubio-Fernández, 2020; Long et al., 2018). It is thus an empirical question as to what underlies pronoun use in Maintenance contexts at different ages. By testing a wide range of ages over the adult lifespan, we can better understand the trajectory of pronominal use at different stages of development (i.e., what happens in between the younger and older years? Are changes gradual or is there a sharp shift in skills at certain age boundaries?). It will also allow us to better investigate the role of age-related cognitive change on referential choice.

Current Study

The aim of the current study is to address these knowledge gaps while expanding on Fossard et al.'s referential complexity scale. Following previous work (e.g., Arnold & Griffin, 2007; Fukumura et al., 2010), we presented participants with two panel scenes at the start of a trial and gave them a sentence describing the first panel; they then were asked to construct a story continuation based on the second panel for a hypothetical addressee.

In each of our experiments, we held the discourse stage constant, allowing us to systematically test the role of referential complexity and aging on referential choice while focusing on an understudied discourse stage: Maintenance. Until recently, the widespread assumption has been that the target referent in Maintenance contexts is always highly topical for both speaker and addressee (Hendriks et al., 2014) and thus should not result in differences in referential choice. However, as mentioned above, emerging work has found that referential choice in Maintenance contexts is not uniform, but rather varies as a function of referential complexity and executive functions (Sandoz et al., 2020, 2023), making it particularly suitable for a study of this nature.

One advantage of studying Maintenance contexts over the commonly studied Re-Introduction stage is that we can compare referential choice across all three levels of Fossard et al.'s referential complexity scale (which is not possible with Re-Introduction as one cannot reintroduce a referent in one-character scenes). Another advantage to studying Maintenance contexts is that the use of pronouns is generally felicitous, providing us with an opportunity to investigate the use of pronouns as discourse markers (Sandoz et al., 2023). Importantly, while pronouns are the preferred marker of topical referents and names the form used to mention less topical referents, the use of names in this context *is still felicitous*. In other words, even though pronouns are optimal for marking topic continuity, a name would still effectively identify the target referent for the listener. As such, contexts like these which permit felicitous variation are well-suited for testing referential complexity effects across different age groups as there are no hard constraints delimiting

[3] Note: "planning skills" may involve a wide range of executive functions, as goal-directed behaviour depends on various cognitive processes that can be recruited in strategic ways. In Experiment 1 we attempt to clarify which type of cognitive "planning skills" are involved in referential choice across the adult lifespan.

what forms are permitted; instead, there are more or less felicitous choices. We thus expect variation between the use of a pronoun and a more explicit referring expression, with the exact choice depending on the context of use (i.e., complexity level) and potential features of the speaker (e.g., their age and cognitive abilities).

Within the Maintenance contexts, we first tested Fossard's referential complexity scale alongside participants' cognitive skills (Experiment 1), then expanded on the scale by examining three novel visual and discourse manipulations as additional factors that could impact referential complexity and lead to age-related differences (Experiments 2–4). Because pronominal use is guided by the likelihood of a given referent as the topic of the discourse (e.g., Lambrecht, 1994; Rohde & Kehler, 2014; Strawson, 1964), we varied the COMPETITION FOR TOPICHOOD between the main character and competitors for each manipulation. Specifically, we varied: (a) the NUMBER of competitors in the scene/discourse (0–2); (b) the TIMING of competitors' presence (early or late in the scene/discourse); and (c) the EMPHASIS on competitors (relative to the main character). This allowed us to pinpoint whether such variations would sway the perceived level of complexity in one direction or another (i.e., up or down the scale; see Figure 1).

In what follows, we present our four experiments, each with an introduction that outlines the theoretically motivated predictions and findings. For the purposes of this research, we were interested in the use of more and less explicit referential forms, thus our coding scheme for each experiment was based on participants' rate of pronominal use relative to names (i.e., 1 = pronouns, 0 = names).

Experiment 1: Testing Age and Executive Control Using Fossard's ET AL (2018) Scale

Our first experiment is a replication of Experiment 1 from Arnold and Griffin (2007) whereby we presented participants ($N = 200$, ages 19–82) with two types of critical trials: one-character scenes (where only one character was present in Panels 1 and 2) and one-character scenes (where two characters of a different sex/gender were present in each panel).

These manipulations correspond to Levels 1 and 2, respectively, on Fossard et al.'s referential complexity scale (see Figure 1).

Results from Arnold and Griffin (2007) revealed that younger adults used pronouns less frequently for two-character than one-character scenes (i.e., for scenes with greater referential complexity), even though a pronoun would have been unambiguous in either condition. In other words, when the sex/gender of the characters is held constant (i.e., unambiguous in either condition), the number of characters appears to influence perceived referential complexity. To date, this effect has been replicated in studies with young adults (e.g., Fukumura et al., 2010), young children (Serratrice, 2013), and more recently older adults (Sandoz et al., 2020, 2023), demonstrating Arnold and Griffin's original findings are robust (2007). However, it has yet to be determined whether adults at different stages of development (young adults vs. middle-aged adults vs. older adults) remain similarly sensitive to this distinction in referential complexity and/or whether general differences in referential behavior will emerge.

As such, this experimental paradigm allowed us to simultaneously test two research questions: (a) In general, do patterns of referential choice vary as a function of age and executive control? (b) More specifically, do patterns of referential choice vary by those factors as well as the referential complexity of the visual scene?

Research Question 1: General referential patterns and cognitive aging.

After an initial analysis to confirm that participants' executive functions varied in the expected direction (i.e., a pattern of decline with increasing age; see Appendix A), we can test the nature of the relationship between adults' overall rate of pronominal use and executive function skills.

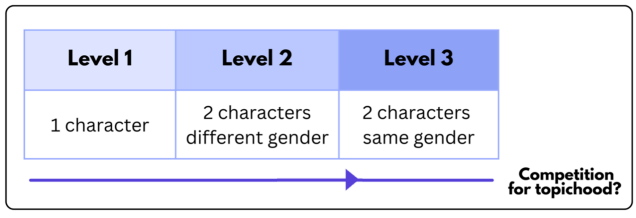
Based on prior work on referential communication, we predicted that younger and older adults would rely on different cognitive strategies relevant to their age and cognitive profile (e.g., Long et al., 2018). Further supporting this prediction are studies that show executive functions follow distinct trajectories over the adult lifespan, with certain skills declining earlier than others (e.g., Daniels et al., 2006; Meiran & Gotler, 2001; Reimers & Maylor, 2005) and at substantially different rates depending on the age group—for example, slow decline in middle age with rapid decline in old age (e.g., Reuter et al., 2019; Veríssimo et al., 2022). As such, it stands to reason older adults would rely on the cognitive resources still available to them to guide referential choice.

The above work highlights a significant point: meaningful distinctions in human behavior are often associated with particular stages of development (e.g., young adulthood, middle age, and old age). As such, it can be advantageous to treat age as both a continuous and categorical variable, to gain a more nuanced picture of the rise and fall of certain skills both collectively and specific to one's age group, a technique often employed in clinical work (e.g., Jorm et al., 2005). Given that our study includes all three adult age groups (including middle-aged adults whose referential behavior has yet to be studied in this context), we will conduct separate continuous and group analyses with regard to age. Specifically, in our first analysis, we will test whether cognitive control predicts pronominal use across the adult lifespan by treating age as a continuous variable. In our second analysis, we will examine the role of cognitive control on referential choice in younger, middle-aged, and older adult groups.

The cognitive measures selected for this study were based on emerging research in aging and referential choice. Specifically, recent work on older adults has found a link between better planning abilities and increased pronominal use as a signal of topic continuity (e.g., Sandoz et al., 2020, 2023). For the purposes of this study, we focused on three critical areas of attentional control (as indices of "planning skills") that have been associated with linguistic behavior

Figure 1

Illustration of Fossard et al.'s (2018) Referential Complexity Scale, With a Question Surrounding the Influence of Our Critical Manipulations



Note. Specifically, do factors relating to COMPETITION FOR TOPICHOOD sway the level of referential complexity (i.e., move the arrow up or down the scale)? See the online article for the color version of this figure.

(Long, Rohde, & Rubio-Fernández, 2020; Long, Vega-Mendoza, et al., 2020; Long et al., 2018): sustained attention, inhibitory control, and attention switching, in addition to working memory.[4]

Inhibition and attention switching play an important role in regulating cognitive control, as outlined in Braver's (2012) "dual-mechanisms of control" framework. According to this framework, the ability to coordinate, regulate, and maintain goal-directed behavior is operationalized through the dynamic use of two semi-independent yet complementary modes of cognitive control: a proactive mode that optimally biases attention to a given goal (an index of inhibitory control) and a reactive mode in which a response is triggered after interference is detected (an index of attention switching). Applying Braver's framework to pronominal use, if your goal as an interlocutor is to prioritize attention to the subject of the sentence and maintain the listener's representation of that active referent, then less attention would be given to the secondary character resulting in a higher rate of pronominal use to refer to the subject. If, however, you prefer a reactive mode of control, then attention would be allocated to each referent mentioned, and as a late corrective mechanism, the secondary character would need to be inhibited while attention is refocused on the subject. While each mode of control involves its own distinct flow of processes, and individuals may favor different modes at different ages, ultimately the same predictions can be made regarding referent accessibility: those with better cognitive control will produce higher rates of pronouns, signaling topic continuity.

Within the cognitive aging literature, there is evidence that older and younger adults favor different modes of cognitive control (e.g., Braver, 2012) which may be driven by the availability of cognitive resources at different ages (Jimura & Braver, 2010; Paxton et al., 2008). While attention-switching skills are more likely to remain intact in old age (Meiran & Gotler, 2001; Reimers & Maylor, 2005; Verhaeghen & Cerella, 2002) continuous sustained inhibition is less likely to remain intact and is arguably more cognitively taxing for older adults (Chiew & Braver, 2017). This would explain why recent work revealed that inhibition was not a robust predictor of pragmatic competence in older adults (Bambini et al., 2021). Based on these findings, we theorize that attention switching will underlie older adults' pronominal use whereas inhibitory control will underlie younger adults' pronominal use. This prediction also raises an interesting question: At what stage of life do potential shifts in cognitive strategies occur? Our separate analyses on younger, middle-aged, and older adults should help to provide an answer.

In addition to attentional functions, working memory may also play a role in discourse planning. Indeed, working memory has been found to influence pragmatic competence across a variety of tasks and ages (e.g., Bambini et al., 2021). As such, it is plausible that those with higher working memory skills are more likely to use pronouns to refer back to the subject referent because they are better able to hold on to and subsequently consult their discourse representation to decide the appropriate referential form to use. If this is the case, we may expect that younger adults are more likely to rely on this type of strategy as older adults tend to experience age-related decline in working memory skills (Craik et al., 1990).

Research Question 2: Testing Fossard et al.'s (2018) scale of referential complexity.

In addressing the second research question, we assessed adults' rate of pronominal use across referential complexity Levels 1 and 2, predicting no age-related differences in referential behavior. This prediction is based on findings from separate studies which have revealed that children, younger adults, and older adults all use pronouns less frequently for two-character than one-character scenes (Arnold & Griffin, 2007; Fossard et al., 2018; Fukumura et al., 2010; Sandoz et al., 2023; Serratrice, 2013), even though a pronoun would be unambiguous in either condition. While it has been theorized that a distinction between one versus two characters may be driven by a division in attentional resources between multiple referential candidates (e.g., Arnold & Griffin, 2007; Karimi, 2022), no study to date has been able to link executive function skills to this two-character effect, even when examining the older adult population whose executive function skills vary dramatically (e.g., Sandoz et al., 2020, 2023).

Based on these previous findings, we expect that adults' sensitivity to the referential complexity of the immediate visual scene to remain stable in these one versus two-character contexts, such that adults of all ages will reduce their pronominal use in two-character scenes. As no study has tested a large sample of younger, middle-aged, and older adults together, our results will shed light on whether this seemingly robust effect replicates at each stage of the adult lifespan.

Given that there is some evidence from studies of pragmatics/reference and aging that including versus excluding fillers differentially affects younger and older adults' referential behavior (Long, Rohde, & Rubio-Fernández, 2020), we decided to test participants in both scenarios. In keeping with Arnold and Griffin (2007), Version 1 of our task included filler trials randomly interspersed in the trial sequence. In Version 2, we removed the filler trials so that a separate set of participants only saw critical trials. If the previously documented filler effect extends to pronominal use, the removal of fillers may reveal age-related differences which could have otherwise been masked. In addition, we expect a general effect of version such that those administered the version with fillers will become more familiarized with the characters resulting in an increase in pronominal use.

Method

Participants

A total of 209 adults were recruited from the University of Edinburgh volunteer panel, the University Careers Services website, and local community groups. All correspondence was done via email, which ensured that everyone who registered was computer literate. Written informed consent was obtained from all participants and the study was approved by the University of Edinburgh Linguistics and English Language ethics committee.

Of the 209 participants who underwent the initial cognitive assessment (detailed in Appendix A), a total of 104 adults went on to participate in Version 1 of the story continuation task (i.e., the version of the task with fillers), while the remaining 105 participants were tested in Version 2 (critical trials only). Participants were assigned to one of the two versions according to their availability for in-lab testing (with the first batch of participants administered Version 1 of the task and the second batch Version 2). Prior to analysis, we excluded four participants from Version 1 (three nonnative

[4] Following previous work, sustained attention was used as a baseline measure to ensure participants did not have severe attentional deficits.

English speakers and one with abnormally low attentional scores) and five participants from Version 2 (three nonnative English speakers, one participant with hearing difficulty, and one due to a technical malfunction), resulting in a total of 200 native English speakers (100 in each group) with normal vision and hearing.

The age ranges in each group were nearly identical: 19–82 in Version 1 and 19–81 in Version 2. However, the age distributions were not parallel in the two groups (see Figure 2). As the median age for each group was different (Table 1), we tested for a main effect of Age across the two versions.[5] A simple linear model (age ~ version) revealed a significant difference in age between the two versions ($p < .001$), such that participants in Version 1 were overall older than those in Version 2. The potential confound of age and version will be addressed in the analyses and interpretation of the results.

Design and Procedure

Story Continuation Task. The story continuation task included 20 critical trials, each consisting of two-panel vignettes featuring one or two animal characters carrying out actions (see Figure 3). A total of 10 possible characters, five male and five female, were randomly combined to create the displays (Males: Birdie, Doggie, Froggy, Horsey, Piggy; Females: Bunny, Duckie, Goosey, Kitty, Mousey). Females were depicted with long eyelashes and usually wore dresses and bows, whereas males wore trousers and ties. Critical trials consisted of one-character vignettes ($n = 10$) and two-character vignettes ($n = 10$). Each participant saw 20 critical trials, with the assignment of one-character or two-character conditions rotated across items in a Latin Square design.

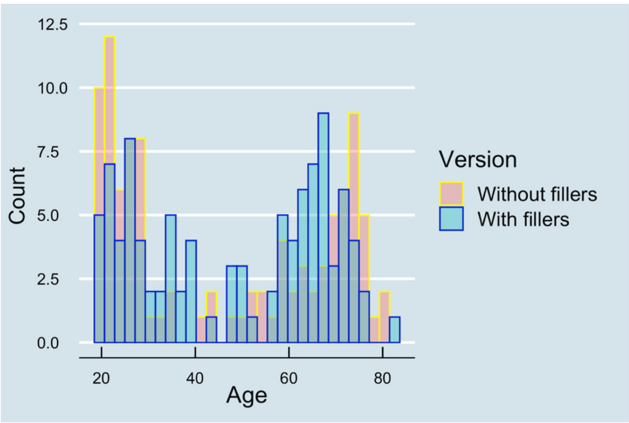
Following Arnold and Griffin (2007), Version 1 of the task included filler trials ($n = 40$) randomly interspersed in the trial sequence. Fillers were a hybrid of one- and two-character trials such that the number of characters varied from Panel 1 to Panel 2, with a transition from 1 → 2 characters ($n = 20$) or 2 → 1 ($n = 20$). The same 10 characters featured in critical trials were also featured in filler trials. Version 2 of the task included the same critical trials, but no filler trials.

Table 1
Model Output for Pronominal Use

Fixed effect	Coefficient	*SE*	*p*
Switching	.7129	.2518	.0046
Working memory	−.0068	.2289	.9763
Inhibition	−.1794	.2165	.4074
Age	.6092	.2189	.0054
Switching × Age	.0127	.2509	.9597
Working Memory × Age	−.2633	.2196	.2304
Age × Version × Complexity	.0159	.2172	.9415

Note. Significant main effects and interactions are shaded.

Figure 2
A Histogram of the Age Counts (i.e., Number of Participants at Each Age) Across Versions of Experiment 1 (i.e., Version 1: With Fillers and Version 2: Without Fillers)



Note. See the online article for the color version of this figure.

At the beginning of each trial, a preview of both panels appeared for 2 s, then the first panel remained in the top half of the computer screen. Participants heard a recorded sentence describing the first panel and they repeated it (for a full list of sentences, see Appendix C); then the second panel appeared beneath the first and participants constructed a story continuation to describe the scene in the second panel. For two-character trials, both characters were depicted as similarly sized in Panel 1. The main character (i.e., the subject of the sentence in Panel 1) was always shown carrying out an action worthy of narration in Panel 2 (e.g., eating rice with a spoon in Figure 3, Panel 2). To further encourage participants to refer to the main character, the secondary character was depicted as relatively inactive and smaller than the main character in Panel 2. To create the one-character trials, we simply removed the secondary character so only the main character was present in the two panels. Main characters were rotated such that each of the 10 characters was featured as the main character in two critical trials. Sentences were presented in the past tense, and when a secondary character was present, the character was mentioned in a "with" phrase (e.g., Doggie cooked rice [with Mousey] for dinner).

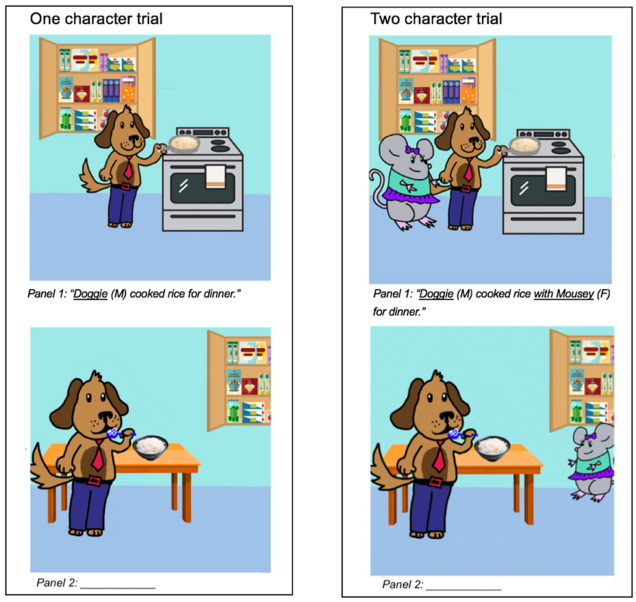
Of interest was the way in which participants referred to the main character (i.e., with a pronoun or a name) in describing Panel 2 and whether the referential form varied depending on the presence or absence of a secondary character in the scene. Importantly, in two-character trials, the characters were always a different sex/gender so a pronoun would distinguish between the two. Thus, the use of a name in this context should reveal the speaker's assessment of the scene as more referentially complex, whether directly or indirectly (e.g., driven by a failure of executive function in the face of more complex scenes), though alternative explanations also exist (e.g., a retrieval error).

At the start of the task, participants were introduced to the 10 animal characters by name and were then asked to identify each of them on their own. To ensure participants of all ages could remember the 10 names, we chose names based on the animal name (e.g., the mouse character was named Mousey). Participants were told to imagine that they were telling stories for a 5-year-old child and to refrain from adding extra humor. After the character familiarization phase, participants were given two practice trials before beginning

[5] We would like to thank Reviewer 2 for inviting us to conduct this analysis, which revealed differences in the age distribution of participants across the two versions of the task.

Figure 3

Sample One- and Two-Character Trials



Note. See the online article for the color version of this figure.

the task. Participants' responses were audio recorded for transcription and coding purposes.

Data Selection Criteria

Practice trials and filler trials were not included in the analysis. Using the same criteria outlined in Arnold and Griffin (2007), we only included responses that referred to the main character as the subject of the sentence and preceded any mention of another character. As such, we excluded cases in which the referring expression denoted more than one character (e.g., *they*, *both of them*, *Doggie and Mousey*, etc.) or was possessive (e.g., *Birdie's ball*, *her ball*, etc.). We also excluded cases of naming errors (for both character and gender), though self-corrections were not excluded. Furthermore, we excluded cases where the referring expression was elliptical (e.g., *…and went to the park*), as well as cases where the response started with a quote from one of the characters (e.g., *Mmm this rice is delicious.*). Finally, we excluded cases where the referring expression was part of a subordinate clause (e.g., *When Goosey* [main character] *dozed off*, *Doggie* [secondary character] *went for a walk*.) unless the referring expression in the subordinate clause did not differ from that in the matrix clause (e.g., *When he* [main character] *got to work*, *he* [main character] *was still feeling ill*.). A total of 199 responses (out of 2,000) were excluded, thus around 10% of the data.

Results

Summary of Initial Cognitive Analyses

A full description of the methods and results from the initial cognitive assessment is presented in Appendix A. Our results revealed a main effect of age on switching ($p < .001$) and working memory ($p < .001$) in the predicted direction: as age increased, switching and working memory decreased. These results provide confidence that there is sufficient age-related variability within our sample. Moreover, the direction of the effect is in line with findings from the aging literature (Gilchrist et al., 2008; Wasylyshyn et al., 2011). While our measure of inhibition showed no effect of age ($p = .196$), this result supports findings from a recent meta-analysis calling into question the notion that inhibition typically deteriorates with age (Rey-Mermet & Gade, 2018). Another possibility for this null effect of age is that our inhibition measure was not sensitive enough. However, as shown in the table and graphs in Appendix A, scores were not uniformly at ceiling (as was found for our baseline measure of sustained attention), thus our participants displayed suitable variability in this measure.

Analyses for Research Question 1: General Referential Patterns

In our first analysis, we tested general referential patterns as a function of age (treating age as a continuous variable) and executive control. We conducted a logistic regression model with the binary outcome variable of pronominal use (pronoun = 1, name = 0) with age, switching, inhibition, and working memory entered as a scaled continuous predictors and maximal random effect structure for participants and items (Barr et al., 2013). All materials, data, and analyses for each experiment are on OSF: https://osf.io/7dzta/?view_only=04a2c177e2564b27a6699cde4ce5bbdd.

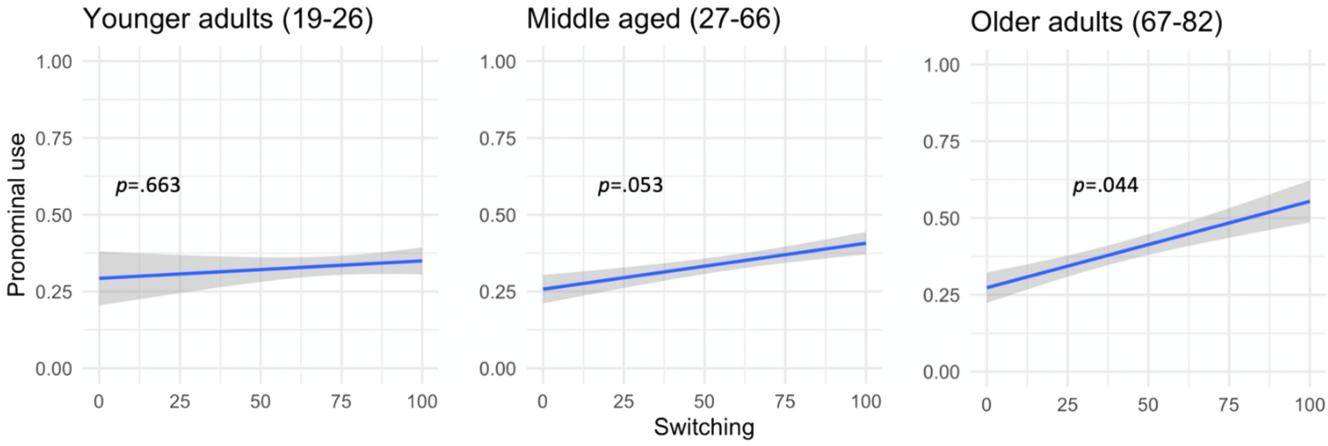
A significant main effect of age emerged ($p = .005$), with greater pronominal use associated with advanced age, a pattern which has been documented extensively in the literature (e.g., Hendriks et al., 2008, 2014). We also found that while inhibition and working memory did not play a significant role in pronominal use ($p = .405$ and $p = .984$, respectively), switching was predictive of pronominal use ($p = .005$), such that better switching was associated with a higher rate of pronouns (see Table 1). This finding helps to clarify which type of cognitive "planning" skills are involved in discourse planning (Sandoz et al., 2023), highlighting the importance of attention switching in guiding the use of optimal discourse markers in Maintenance contexts (i.e., pronouns for topic continuity).

Our second planned analyses treated age as a categorical variable, allowing us to gain a deeper understanding of both the relationship and trajectory of switching and pronominal use at different stages of development. To do so, we performed a tertile age split, dividing subjects into three equal groups: those in younger adulthood (19–26 years), middle adulthood (27–66 years), and older adulthood (67–82 years). The age range for the younger and older groups is similar to those from prior work (e.g., Hendriks et al., 2008, 2014), allowing us to compare our results with previous findings, while at the same time studying the in-between years (i.e., the middle stage of adulthood that has yet to be investigated in this context).

For each age group, we modeled the binary outcome variable of pronominal use (pronoun = 1, name = 0) with switching entered as a scaled continuous predictor and maximal random effect structure for participants and items (Barr et al., 2013). Our results revealed that older adults (and to a lesser extent those in the middle-aged group) appear to be driving the effect of switching (Figure 4), likely because older adults whose switching capacity is still preserved rely heavily on that ability to guide effective communication (Braver, 2012). Following a steady developmental trend, younger

Figure 4

Plots of the Tertile Age Split Showing the Relationship Between Switching and Pronoun Use



Note. See the online article for the color version of this figure.

adults' referential behavior is not determined by switching skills ($p = .663$), middle-aged adults' referential behavior begins to show a marginal (though nonsignificant) relationship ($p = .053$), while older adults' referential behavior shows a significant relationship ($p = .044$). These analyses demonstrate the importance of studying adults at each stage of development (not just extremes, such as younger vs. older adults) as we can gain a clearer picture of the stage in which adults shift to a reliance on a specific set of cognitive functions to guide referential choice. Together, our two analyses suggest that there is a trend toward an increased reliance on switching for referential choice as we age, and that this is driven by older adults (and to a lesser extent middle-aged adults) whose use of pronouns is strongly linked to switching skills.

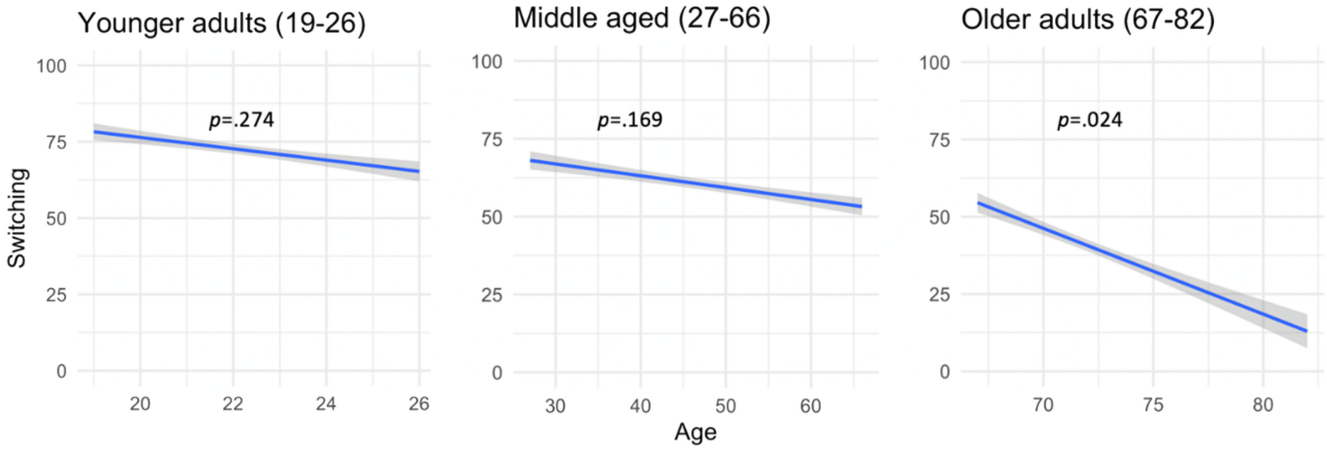
Our finding that older adults' switching skills underlie pronominal use mirrors results from the pragmatics and aging literature (Long et al., 2018; Long, Rohde, & Rubio-Fernández, 2020) and cognitive aging literature (e.g., Braver, 2012), whereby older adults seem to rely more on a reactive mode of cognitive control (involving attention switching) than a proactive mode of control (involving inhibition). This could be due to a compensatory shift in older adults from a proactive to a reactive strategy in the face of age-related neurocognitive changes, as previously documented (e.g., Braver, 2012; Jimura & Braver, 2010; Paxton et al., 2008) or to greater variability in switching in older adulthood, making it a stronger predictor of pronominal use in that population. Either way, it is noteworthy that switching is what underlies older adults' referential behavior considering that many older adults showed decline in that area (see Appendix A).

For a comparative analysis, we reanalyzed the switching data using the tertile age split. As shown in Figure 5, a similar linear trajectory emerges as in Figure 4, but this time in the opposite direction, with a steady decline in switching skills that intensifies in the older group.

When comparing the tertile age findings, it may seem counterintuitive that older adults would rely on switching skills for discourse

Figure 5

Plots of the Tertile Age Split Showing the Relationship Between Age and Switching Skills



Note. See the online article for the color version of this figure.

planning when those skills are in decline. However, we interpret our results to suggest that older adults whose switching skills are still in-tact recruit those skills for effective communication. This interpretation is supported by studies in pragmatics and aging which show a similar trend toward a decline in switching with advanced age alongside an increased reliance on switching (Long, Rohde, & Rubio-Fernández, 2020; Long et al., 2018).

Taken together, our findings suggest that in Maintenance contexts an increase in pronominal use may be partially driven by a reliance on switching skills to guide discourse planning in old age. This interpretation is in line with the explanation provided by Sandoz et al. in their studies with older adults (2020, 2023) whereby enhanced "planning skills" (in this case switching) can be understood to reflect a participant's capacity to guide their addressee to maintain focus on the intended referent. Following this logic, if older adults are more likely to prioritize marking topic continuity for the listener through pronominal use (even though maintaining a referent's activation may be cognitively costly), they may rely heavily on executive control to do so. On the other hand, younger adults (and to a lesser extent middle-aged adults), who have been found to use far less pronouns for maintenance (Hendriks et al., 2014) may instead prioritize the use of more explicit referential forms as a default communicative heuristic based on the notion that explicitness removes any potential for referential ambiguity (Note that this is one possible interpretation of the data; see "General Discussion" for a more in-depth discussion).

Main Analysis

To test the combined effects of complexity, version, and age on referential choice, we modeled the binary outcome variable of pronominal use (pronoun = 1, name = 0) with complexity (1 → 1 or 2 → 2/different gender), age, and version (with or without fillers) and their interactions as predictor variables. Age was entered as a scaled continuous predictor and deviation coding was used for complexity (1 → 1 = −0.5, 2 → 2/different gender = 0.5) and version (with fillers = −0.5, without fillers = 0.5). The model was fit with the maximal random effect structure for participants and items (Barr et al., 2013).

Descriptive statistics for participants' performance on each version of the task are reported in Table 2. Below we review the impact of the manipulated conditions on participants' referential choice, but first, we note the lack of a main effect or interaction with age (all $ps > .05$). The absence of interactions suggests that neither version nor complexity was affected by age-related changes to cognitive control[6]; instead, as predicted, adults of all ages display the same patterns of referential behavior, demonstrating that they are equally sensitive to those task manipulations. Regarding the absence of a main effect of age (which was significant in the previous, smaller model for the combined data), we conclude that it is likely masked here by the multiple two- and three-way interactions which were not significant (see Hawkins, 2004 on model overfitting).[7]

In line with previous work, our results revealed a main effect of complexity, in which participants of all ages used more pronouns (vs. names) in 1 → 1 than 2 → 2/different-gender scenes ($p < .001$; for full model output, see Table 3). As predicted, we found a main effect of version ($p = .0319$), with a higher rate of pronominal use in the version with fillers than in the version without fillers (Figure 6). One way of interpreting this finding is that participants may have become more familiarized with the characters and were thus more likely to opt for a pronoun over a proper name. Supporting this interpretation is the finding that there was an increased rate of pronominal use during the second half of the task ($m = .43$) compared to the first half ($m = .37$; $p = .0254$), which was not the case for the version of the task without fillers, where no significant difference was found between blocks ($p = .145$).

While the above interpretation is supported by our data, the confound of age cannot be excluded as a possible explanation for the main effect of the version, nor can the difference in the number of trials (with a larger number of trials in the version with fillers). In other words, it is unclear whether the effect of version was brought about by an increase in familiarity with the characters in Version 1, by the older age of participants in Version 1, or by the difference in task length. In Experiments 2–4, we eliminate potential confounds by removing filler trials altogether and recruiting a more balanced sample in each experiment, whereby the distribution of ages is even across 10-year age bands.

Table 2
Descriptive Statistics for Experiment 1

One character	*M*	*SD*	Range
Version 1 (with fillers)			
All participants	.54	.50	0–1
Younger adults	.45	.50	0–1
Middle-aged	.54	.50	0–1
Older adults	.63	.48	0–1
Two character	***M***	***SD***	**Range**
All participants	.25	.43	0–1
Younger adults	.22	.42	0–1
Middle-aged	.21	.41	0–1
Older adults	.33	47	0–1
One character	***M***	***SD***	**Range**
Version 2 (without fillers)			
All participants	.47	.50	0–1
Younger adults	.54	.50	0–1
Middle-aged	.42	.49	0–1
Older adults	.48	.50	0–1
Two character	***M***	***SD***	**Range**
All participants	.13	.34	0–1
Younger adults	.16	.37	0–1
Middle-aged	.12	.33	0–1
Older adults	.11	.32	0–1

Note. The table shows the rate of pronominal use for all participants as well as a tertile age split for each version of the task. For the purposes of this table, we carried out tertile age splits. For Version 1, younger adults were between the ages of 19 and 27, middle-aged between 28 and 65, and older adults between 66 and 82. For Version 2, younger adults were between the ages of 19 and 26, middle-aged between 27 and 66, and older adults between 69 and 81.

[6] In line with the previous analyses, we also ran a model with the three cognitive measures as additional factors. As expected, none of the cognitive measures interacted with Complexity to predict pronominal use (see Table B1 in Appendix B for full model output).

[7] We would like to thank Reviewer 1 for suggesting that we initially run smaller models for each experiment to test the effect of Age on pronominal use, which as they anticipated, revealed clearer results than the larger models with multiple interactions.

Table 3
Model Output for Pronominal Use From the Combined Data

Fixed effect	Coefficient	*SE*	*p*
Complexity	−3.2404	.3191	<.001
Age	.4314	.2631	.1011
Version	−1.1538	.5377	.0319
Age × Complexity	−.3148	.2651	.2351
Version × Complexity	−.5937	.5550	.2847
Age × Version	−.8615	.5237	.1000
Age × Version × Complexity	.1556	.5157	.7629

Note. Significant main effects and interactions are shaded.

Experiment 2: Extending the Referential Complexity Scale: NUMBER of Competitors

In our second experiment, we tested whether an increase in the NUMBER of competitors in the scene/discourse (from 0 to 2) could impact referential complexity due to COMPETITION FOR TOPICHOOD. Using the same critical trials from Experiment 1, we increased the NUMBER of competitors in the scene/discourse to test whether complexity increases when there are two competitors of a different sex/gender than the main character in the immediate visual scene.

Fossard et al. (2018) did not place these types of three-character scenes within their referential complexity scale (i.e., as low-, intermediate-, or high-level complexity) or specify what effect these scenes should have on adults' referential choice. However, Fossard et al.'s scale distinguish between one- and two-character scenes as increasing in referential complexity, and adults of all ages adapted their referential choices according to this discourse manipulation in the previous experiment. Therefore, following Fossard et al.'s distinction between one- and two-character scenes and the results of Experiment 1, we were interested in whether three-character scenes (specifically those in which two competitors of a different sex/gender are competing with the main character for TOPICHOOD) would result in a further reduction in pronominal use (i.e., one character > two characters > three characters), indicating an increase in referential complexity. We were also interested in whether this manipulation would reveal age-related differences in referential choice.

Method

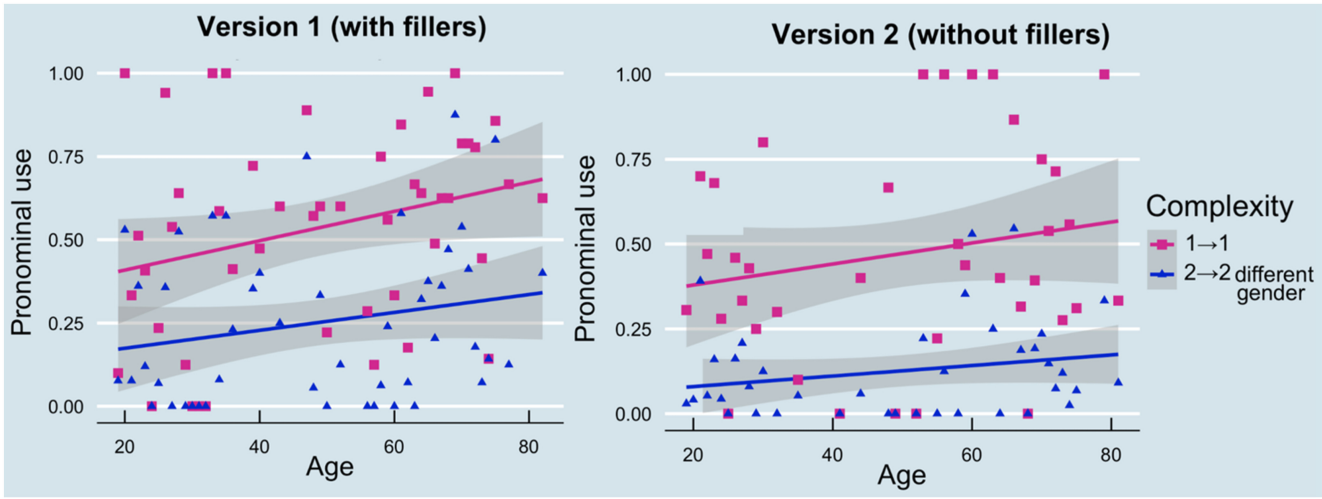
Participants

Due to the ongoing global pandemic, we were unable to continue testing in-person and instead recruited participants for web-based testing (which entailed a similar experimental setup but with written rather than oral responses and no cognitive measures). For this experiment, we recruited 100 Scottish participants from Prolific, an online crowdsourcing platform. To ensure that we had a similar range of ages as in the previous experiment as well as an even spread of ages, we recruited a total of 100 participants, 20 participants in each of the following age bands: 18–28, 29–39, 40–50, 51–61, and 62 and above. Prior to commencing the task, informed consent was obtained from all participants. The study was approved by the University of Edinburgh Linguistics and English Language ethics committee. Before analysis, data from four participants were removed as all of their responses were incomplete (e.g., one-word answers). Thus, we report data from 96 participants ages 18–73 who confirmed that they had normal vision and hearing.

Design and Procedure

The same critical items from Experiment 1 were used in Experiment 2. In addition to these 20 items, 10 new items were added to provide equivalent power for the three-level condition of number of characters (for a full list of sentences, see Appendix C). Each participant saw 30 critical items, with the assignment of 1 → 1, 2 → 2, or 3 → 3 character conditions rotated across items in a

Figure 6
Pronominal Use by Age and Version for 1 → 1 and 2 → 2/Different Gender in Experiment 1



Note. Regression lines reflect the best fit of data, points reflect mean pronominal use for each age tested. The shaded bands around the regression lines represent a 95% confidence region for the regression fit. See the online article for the color version of this figure.

Latin Square design. When a third character was present, it was always the same sex/gender as the secondary character, which differed from that of the main character (e.g., if the main character was female, the two other characters were male). This was to ensure that the use of a pronoun would distinguish the main character from the other characters. For the three-character scenes, all characters were depicted as being similarly sized in the first panel. In the second panel, the second and third characters were depicted as smaller and less active than the main character (see Figure 7). As with the previous experiment, this depiction was used to encourage participants to refer to the main character in their story continuation. In the opening sentences for the three-character scenes, the third character was always mentioned in an "and also" phrase (e.g., Bunny went to the café with Mousey [and also Piggy].) The inclusion of "also" in the conjunction was intended to highlight the presence of the third character in the text. Since we were interested in testing whether participants treated two- and three-character scenes differently, we wanted to ensure that the two- and three-character sentences were clearly distinguishable from one another without making drastic changes to the sentence structure. The same sentences from Experiment 1 were used in Experiment 2, with one small adjustment: to ensure that the three-character sentences sounded natural, we moved the phrase referring to the additional character(s) to the end of the sentence (e.g., instead of "Doggie cooked rice [with Kitty] [and also Piggy] for dinner," we used "Doggie cooked rice for dinner [with Kitty] [and also Piggy]"). In line with the previous experiment, participants were told to imagine that they were telling stories for a 5-year-old child and to refrain from adding extra humor.

Results

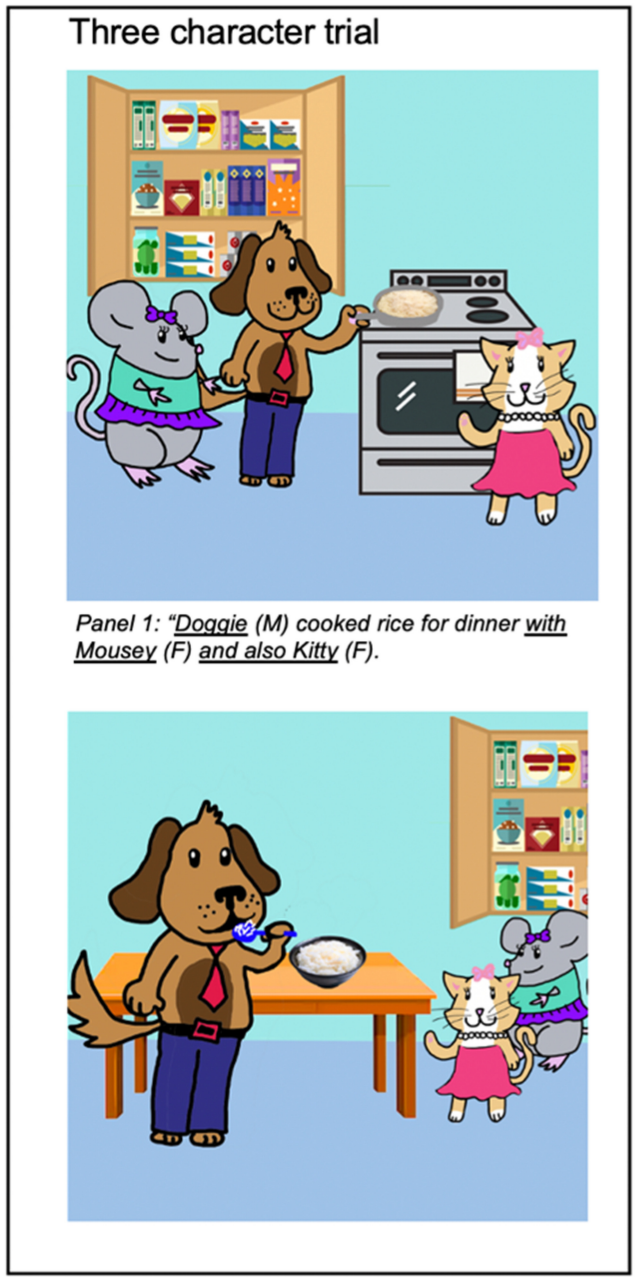
Initial Analysis: General Referential Patterns Across Adulthood

For our initial analysis, we again conducted a logistic regression model in which we modeled the binary outcome variable of pronominal use (pronoun = 1, name = 0) with age (entered as a scaled continuous predictor). This allowed us to test our main prediction that older adults would produce more pronouns (vs. names) than younger adults. In line with the results from Experiment 1, and previous work (e.g., Hendriks et al., 2008, 2014) there was an overall effect of age on referential choice ($p = .0016$), such that older adults were more likely to use pronouns than younger adults.

Main Analysis: The Combined Effects of Complexity and Age on Referential Choice

To test the combined effects of complexity and age on referential choice, we used logistic mixed effects regression to model the binary outcome variable of pronominal use (pronoun = 1, name = 0) with complexity (1 → 1, 2 → 2, 3 → 3), age, and their interactions as predictor variables. Age was entered as a scaled continuous predictor. The model was fit with the maximal random effect structure for participants and items. Complexity was dummy coded with 2 → 2 as the reference level. This reference level was chosen to allow us to determine whether the effect from Experiment 1 (i.e., 1 → 1 vs. 2 → 2 characters) would replicate in the web-based paradigm, and whether participants treated two-character and three-character scenes (i.e., those with one vs. two competitors) differently, since according to Fossard et al. (2018) referential complexity should increase according to the number of characters (or competitors) in the scene/discourse.

Figure 7
Sample Three-Character Trial



Note. See the online article for the color version of this figure.

Results

Descriptive statistics for participants' performance in each condition are reported in Table 4. Before reviewing the manipulated conditions, we again note the lack of an effect or interaction with age (all $ps > .05$). As before, we attribute the absence of a main effect of age in this larger model to the inclusion of multiple

Table 4
Descriptive Statistics for Experiment 2

Participants	*M*	*SD*	Range
One character			
All participants	.52	.50	0–1
Younger adults	.49	.50	0–1
Middle-aged	.49	.50	0–1
Older adults	.61	.49	0–1
Two character			
All participants	.24	.43	0–1
Younger adults	.19	.39	0–1
Middle-aged	.22	.41	0–1
Older adults	.35	.48	0–1
Three character			
All participants	.24	.43	0–1
Younger adults	.19	.40	0–1
Middle-aged	.26	.44	0–1
Older adults	.25	.44	0–1

Note. The table shows the rate of pronominal use in each condition for all participants as well as by age group (determined by a tertile age split). For the purposes of this table, we carried out a tertile age split of the data such that younger adults were between the ages of 18–31, middle-aged between 32 and 58, and older adults between 59 and 73.

Table 5
Model Output for Pronominal Use in Experiment 2

Fixed effect	Coefficient	*SE*	*p*
One-character	2.6121	.3557	<.001
Three-character	.4232	.3657	.2470
Age	.3682	.3556	.3000
One-Character × Age	−.1756	.2732	.520
Three-Character × Age	−.2340	.2152	.2770

Note. Significant main effects and interactions are shaded. The reference level is the two-character condition.

interactions which were not significant, which in turn is likely to mask the effect of age.

Our results replicated Arnold and Griffin's findings as well as the results from Experiment 1: there was an effect of complexity ($p < .001$) whereby pronominal use (vs. names) was greater in the 1 → 1 condition than the 2 → 2 character condition (for full model output, see Table 5). We found no difference between 2 → 2 and 3 → 3 character scenes, with participants using pronouns to refer to the main character at similar rates when there were one or two competitors in the scene/discourse. As increased complexity should lead to more explicit referring expressions (i.e., names; Fossard et al., 2018), our finding suggests that the addition of a second competitor of a different sex/gender does not increase referential complexity, or at least not to the extent that it influences referential choice.

As mentioned above, no other effects or interactions were found, including no interaction of age and condition (see Figure 8). In line with the results of Experiment 1, it appears that manipulation of referential complexity via number of competitors does not result in age-related differences in referential choice.

Experiment 3: Extending the Referential Complexity Scale: TIMING of Competitors

In our third experiment, we tested whether the TIMING of competitors' presence in the immediate visual scene would impact referential complexity (due to early or late COMPETITION FOR TOPICHOOD) and whether that would in turn lead to age-related differences in referential choice. To do so, we created a hybrid manipulation between levels one and two of the referential complexity scale which allowed us to test whether varying the visual presence of the competitor in the first or second panel would influence referential choice (i.e., 1 → 2/different vs. 2 → 1/different).

Prior work on the role of timing in young adults' referential choice has examined the following conditions: 1 → 1, 2 → 2/different gender, and 2 → 1/different gender (Arnold & Griffin, 2007). There the rate of pronominal use was found to be similar for 2 → 2 and 2 → 1, but differed from 1 → 1 trials, where pronominal use was significantly higher. The results of Arnold and Griffin (2007) therefore suggest that the key discourse feature is the number of characters in the first panel—rather than the number of characters in the second panel, which participants were describing. Here we test whether middle-aged and older adults also make early calculations of TOPICHOOD (as indexed by a lower rate of pronouns in 2 → 1 vs. 1 → 2) or whether their calculations come later, based solely on the immediate visual context in Panel 2.

In addition to this, we tested contexts with a higher level of complexity (i.e., level three: where the scene/discourse involves another character of the same sex/gender as the main character). In Arnold and Griffin's (2007) pilot study, they found that pronominal use decreased when the sex/gender of the characters was the same (vs. different). Interestingly, however, recent work on younger adults (Fossard et al., 2018) and older adults (Sandoz et al., 2020, 2023) has not replicated this pattern. Instead, participants in those studies distinguished between one- and two-character scenes, but not between two-character scenes with the same- versus different-gender characters. Given this inconsistency in results, we added a third condition to our experiment: 2 → 2 same gender. This allowed us to test both the role of TIMING of competitors as well as gender ambiguity on referential choice over the adult lifespan.

In sum, we compared 1 → 2/different gender, 2 → 1/different gender, and 2 → 2/same gender. The comparison of the first two allows us to test the role of TIMING; the comparison of 2 → 2/same with either of the other two conditions allows us to test the role of gender ambiguity in contexts with two referents.

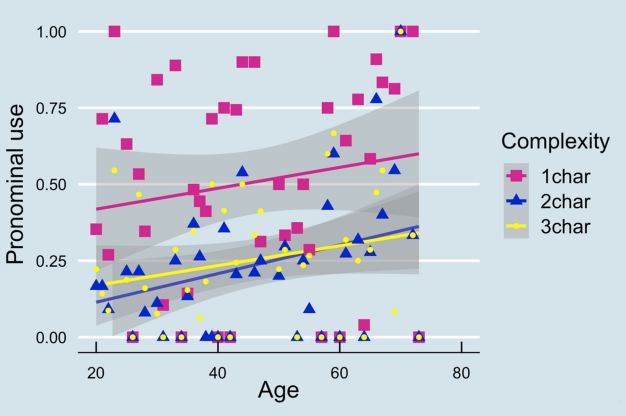
Method

Participants

A new group of 100 Scottish participants ages 19–77 were recruited via Prolific. To ensure an even distribution of ages like in the previous experiments, 20 participants were recruited from each of the following age bands: 18–28, 29–39, 40–50, 51–61, and 62 and above. Prior to analysis, we excluded and replaced four participants: two participants whose responses were always elliptical constructions (e.g., *…and went outside*), one participant whose continuations were all very short and did not mention a referent (e.g., *great time* or *fell asleep*), and one participant who only responded with quoted speech (e.g., *this is so much fun*). Informed consent was obtained from all participants and the study was approved by the University of Edinburgh Linguistics and English Language ethics committee.

Figure 8

Pronominal Use by Age for One, Two, and Three-Character Scenes in Experiment 2



Note. Regression lines reflect the best fit of data, points reflect mean pronominal use for each age tested. The shaded bands around the regression lines represent a 95% confidence region for the regression fit. See the online article for the color version of this figure.

Design and Procedure

Our task included 30 trials with the same experimental setup from Experiments 1 and 2. For 10 of the trials, the main character appeared alone in the first panel, and the main character and a secondary character of a different sex/gender in the second panel (i.e., 1 → 2/different gender). Another 10 trials consisted of the main character and a secondary character of a different sex/gender in the first panel and the main character alone in the second panel (thus 2 → 1/different gender). The other 10 trials consisted of a main character and secondary character of the same sex/gender in each panel (i.e., 2 → 2/same gender; see Figure 9 for a sample illustration). Each participant saw the 30 items in randomized order. The manipulation of complexity was within subjects and between items.

Results

Initial Analysis: General Referential Patterns Across Adulthood

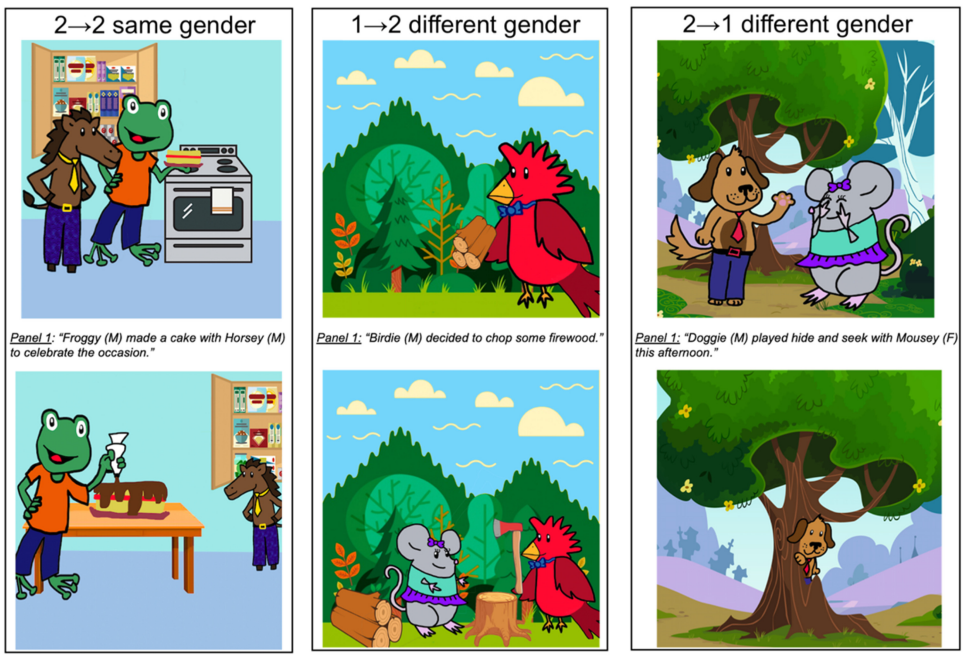
As we were again interested in whether we would find age-related differences in overall patterns of referential choice, we conducted a logistic regression model in which we modeled the binary outcome variable of pronominal use (pronoun = 1, name = 0) with age (entered as a scaled continuous predictor). As predicted, there was a significant effect of age on overall pronominal rates ($p = .0023$) such that older adults used pronouns more than younger adults, in line with prior results from the literature (e.g., Hendriks et al., 2008, 2014).

Main Analysis: The Combined Effects of Complexity and Age on Referential Choice

To test the effects of complexity and age on referential choice, we used logistic mixed effects regression (with maximal random effect structure for participants and items; Barr et al., 2013), modeling the binary outcome variable of pronominal use (pronoun = 1, name = 0) with age, and complexity (2 → 2/same gender, 1 → 2/different

Figure 9

Sample 2 → 2/Same Gender, 1 → 2/Different Gender, and 2 → 1/Different Gender Trials



Note. See the online article for the color version of this figure.

gender, and 2 → 1/different gender) and their interaction as predictor variables. Age was entered as a scaled continuous predictor. Complexity was dummy coded with 2 → 1/different gender as the reference level. This reference level allowed us to test (a) the TIMING of competitors' presence (2 → 1 vs. 1 → 2/different gender) and (b) the role of gender (2 → 1/different vs. 2 → 2/same) on referential choice across the adult lifespan.

Descriptive statistics for participants' performance in each condition are reported in Table 6. Similar to the previous pair of experiments, we first note that there were no main effects of age (all $ps > .05$), which we believe may have again been masked, potentially due to overfitting of the model.

Here our results revealed a difference between 2 → 1/different-gender and 1 → 2/different-gender trials ($p < .001$), such that participants used more pronouns in the latter condition (see Figure 9; for full model output, see Table 7). This supports the notion that participants make complexity calculations early, based on the first scene/discourse, by decreasing their use of pronouns when a competitor is present. Notably, there was also a 1 → 2/different gender by age interaction relative to 2 → 1/different gender ($p = .021$). Follow-up analyses revealed that this interaction was driven by an effect of age for 1 → 2/different-gender trials ($p = .044$) but not 2 → 1/different-gender trials ($p = .962$), whereby pronominal use for 1 → 2/different-gender trials increased with advanced age (Figure 10). Together, these findings highlight two critical points: (a) adults across the lifespan make early calculations of prominence based on COMPETITION FOR TOPICHOOD in Panel 1, therefore the "timing" of competitors' presence should be integrated into Fossard et al.'s referential complexity scale and (b) older adults are more likely than younger adults to prioritize maintaining focus on the referent in initial one-character scenes.

Regarding the difference between 2 → 1/different-gender and 2 → 2/same-gender trials, no difference was found ($p = .178$). In terms of the *number* of characters, these results replicate Arnold and Griffin's findings whereby younger adults' pronominal use is based on the number of characters in the first panel (and its corresponding sentence) rather than the number of characters in the second panel (which participants were describing) and extend these results to middle-aged and older adults. This would explain why the number of characters in the scene/discourse only affects referential choice when comparing 2 → 2 versus 1 → 2 trials and not 2 → 2 versus 2 → 1 trials.

Our findings also support the notion that the presence of multiple characters from the outset of the narrative has a bigger impact on referential choice than gender ambiguity (since 2 → 2/same gender and 2 → 1/different gender have the same initial number of characters and the conditions did not yield different pronoun rates). These findings go against the results from Arnold and Griffin's (2007) pilot study but align with findings from numerous recent studies on younger and older adults' referential choices in these contexts (Fossard et al., 2018; Sandoz et al., 2020, 2023). One explanation is that it is less likely that a pronoun will be perceived as ambiguous in Maintenance contexts since the referent was mentioned in the subject position immediately prior. This interpretation should be taken with caution, however, since the number of characters was not held constant across conditions—an important manipulation that will be tested in the subsequent experiment. Regardless, it is noteworthy that older adults' sensitivity to referential complexity and overall referential behavior does not decline with age, which suggests that in Maintenance contexts, referential complexity is evaluated in a relatively consistent manner across the adult lifespan.

Table 6
Descriptive Statistics for Experiment 3

Participants	*M*	*SD*	Range
2 → 2 same gender			
All participants	.19	.39	0–1
Younger adults	.16	.36	0–1
Middle-aged	.20	.40	0–1
Older adults	.21	.41	0–1
1 → 2 different gender			
All participants	.67	.50	0–1
Younger adults	. 62	. 49	0–1
Middle-aged	.67	.47	0–1
Older adults	.73	.44	0–1
2 → 1 different gender			
All participants	.20	.40	0–1
Younger adults	.20	.40	0–1
Middle-aged	.19	.39	0–1
Older adults	.22	.42	0–1

Note. The table shows rate of pronominal use in each condition for all participants as well as by age group (determined by a tertile age split). For the purposes of this table, we carried out a tertile age split of the data such that younger adults were between the ages of 19–30, middle-aged between 31 and 55, and older adults between 56 and 77.

Table 7
Model Output for Pronominal Use in Experiment 3

Fixed effect	Coefficient	*SE*	*p*
Age	.1323	.2840	.6414
1 → 2 different gender	4.0273	.5304	<.001
2 → 2 same gender	−.6318	.4695	.1784
1 → 2 Different Gender × Age	.9991	.4314	.0206
2 → 2 Same Gender × Age	.3995	.2323	.0855

Note. Significant main effects and interactions are shaded. The reference level is 2 → 1/different.

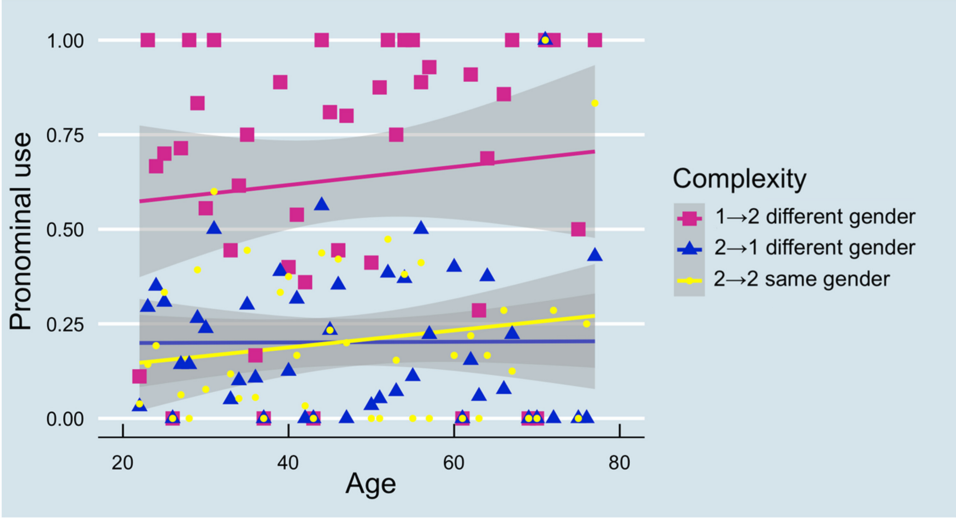
Experiment 4: Extending the Referential Complexity Scale: EMPHASIS on Competitors

Our final experiment tested whether a decreased EMPHASIS on competitors (through re-mentioning of the main character) would impact referential complexity and lead to age-related differences in referential choice. In theory, the additional use of a name or pronoun to refer to the main character should decrease COMPETITION FOR TOPICHOOD between characters by lessening the likelihood of the competitor as the topic, therein increasing the main characters' privileged status (Karimi, 2022). In turn, this should lead to an increase in pronominal use (e.g., Rohde & Kehler, 2014). To what extent this prominence manipulation plays a role in referential complexity or will lead to age-related differences is unclear.

To test this, we used the same conditions from Experiment 3 (i.e., 1 → 2/different gender, 2 → 1/different gender, and 2 → 2/same gender), adding an additional sentence to each prompt that either (a) repeated the main character's name or (b) used an additional pronoun to refer back to that referent. The latter condition is expected to increase the main character's prominence even more as pronouns are generally reserved for highly topical characters (e.g., Rohde &

Figure 10

Pronominal Use by Age for 1 → 2/Different Gender, 2 → 1/Different Gender, and 2 → 2/Same Gender Scenes in Experiment 3



Note. Regression lines reflect the best fit of data, points reflect mean pronominal use for each age tested. The shaded bands around the regression lines represent a 95% confidence region for the regression fit. See the online article for the color version of this figure.

Kehler, 2014), and this may in turn lead to an increase in participants' use of pronouns, corresponding to a decrease in referential complexity.

In addition, we added a 2 → 2/different-gender condition, allowing us to test complexity at two levels on Fossard's scale: intermediate-level complexity (two characters of a different gender) and high-level complexity (two characters of the same gender) in younger, middle-aged, and older adults. In sum, we tested four conditions: 1 → 2/different gender, 2 → 1/different gender, 2 → 2/different gender, and 2 → 2 same gender, and manipulated the prominence of the main character in each condition through one of two types of repeated mention: either an additional name or an additional pronoun.

Method

Participants

A new group of 100 Scottish participants ages 18–73 were recruited via Prolific. Again, in order to ensure an even distribution of ages, 20 participants were recruited from each of the following age bands: 18–28, 29–39, 40–50, 51–61, and 62 and above. All participants met the inclusionary criteria therefore none were removed. Informed consent was obtained from all participants and the study was approved by the University of Edinburgh Linguistics and English Language ethics committee.

Design and Procedure

The same materials and procedure from Experiment 3 were adapted for use in Experiment 4 (for a full list of sentences, see Appendix C). Here a total of four conditions were tested (1 → 2/different gender, 2 → 1/different gender, 2 → 2/different gender, and 2 → 2 same gender). Again, the manipulation of complexity (which here encompassed both EMPHASIS and gender/number manipulations) was within subjects and between items.

To test whether the results from Experiment 3 would replicate when holding the number of characters constant in both panels, we added 2 → 2/different-gender trials ($n = 10$). In addition, we introduced two linguistic manipulations intended to emphasize the main character's prominence which should result in an increase in the use of pronouns over names. First, since overall pronominal use in Experiment 3 was relatively low compared to previous studies (e.g., Arnold & Griffin, 2007; Fukumura et al., 2010), we emphasized the main character's prominence by adding a second sentence to the first panel in 1 → 2 and 2 → 1/different-gender trials. In the second sentence, a pronoun was used in the subject position to refer to the main character (e.g., *Kitty went to play in the back garden. She was in a great mood*). To ensure we had enough power for this new manipulation, we doubled the number of 1 → 2 and 2 → 1 trials.

Second, for the 2 → 2 same- and different-gender trials, we increased the prominence of the main character by repeating its name. The use of a repeated name (vs. pronoun) allowed us to avoid potential ambiguity regarding which character was re-mentioned. To do so, we added a second sentence to the first panel which referenced the main character in a nonsubject position (e.g., *Doggie cooked rice with Mousey for dinner. Rice was Doggie's favorite meal*).

Results

Initial Analysis: General Referential Patterns Across Adulthood

Following the previous experiments, we again tested whether older adults produced more pronouns overall than younger adults,

by conducting a logistic regression model in which we modeled the binary outcome variable of pronominal use (pronoun = 1, name = 0) with age (entered as a scaled continuous predictor).

The results revealed an effect of age on overall rates of pronominal use ($p < .001$). However, unlike the previous experiments, the effect did not go in the expected direction. Instead, younger adults showed a higher rate of pronominal use than older adults. What may be driving this effect? Experiments 1–3 involved manipulations that were both visual and linguistic, whereas Experiment 4 manipulations were linguistic only (as no changes were made to the immediate visual context). A possible explanation for the reverse pattern of results is that older adults are less sensitive to linguistic-only cues than younger adults when calculating a referent's status.

Upon further investigation, it seems that younger adults indeed may have a stronger response to the prominence manipulations than older adults. In Experiments 1–3, younger adults' average rate of pronominal use was .31, whereas in Experiment 4 it increased to .58. Older adults, on the other hand, averaged .40 in Experiments 1–3, with a smaller increase to .47 in Experiment 4. While this observation should be interpreted with caution (as there were different participants in each study), it highlights the possibility of an increased linguistic sensitivity in younger adults which should be further explored in future work.

Main Analysis: The Combined Effects of Complexity and Age on Referential Choice

To test the role of EMPHASIS and gender/number manipulations on pronominal use over the adult lifespan, we used logistic mixed effects regression, modeling the binary outcome variable of pronominal use in critical trials (pronoun = 1, name = 0) with age (entered as a scaled continuous predictor) and complexity (2 → 2/different gender, 2 → 2/same gender, 1 → 2/different gender, 2 → 1/different gender) and their interaction as predictor variables. The model was fit with the maximal random effect structure for participants and items. Complexity was dummy coded with 2 → 2/different gender as the reference level. This allowed us to compare 2 → 2/different gender to 2 → 2/same gender (like in Arnold and Griffin's pilot experiment) as well as 2 → 2/different gender to 2 → 1/different gender (like in Arnold and Griffin's second experiment).

Descriptive statistics for participants' performance across conditions are reported in Table 8. Note that again, as with the previous experiments, we found no main effects or interactions with Age (all $ps > .05$). The absence of a main effect of age in this larger model was likely caused by model overfitting, with multiple interactions which were not significant.

Our results revealed that pronominal use for both 1 → 2 and 2 → 1/different-gender trials was greater than for 2 → 2/different-gender trials (for full model output, see Table 9). This result contrasts with the findings from Arnold and Griffin whereby pronominal use was similar for 2 → 2 and 2 → 1/different-gender trials. Unlike Arnold and Griffin's task, however, ours emphasized the prominence of the main character in different ways. For 2 → 2 trials, prominence was emphasized using a repeated name (thus boosting the main character's prominence through frequency of mention), whereas for the 1 → 2 and 2 → 1 trials, an additional pronoun was used (thus boosting prominence through both frequency of mention and pronominalization, the latter likely signaling that a referent is still the topic; Arnold, 1998; Arnold et al., 2009; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). The use of two prominence manipulations may have further increased participants' likelihood of continuing the story with another pronoun. It is also possible that the use of a name in the other condition did not have the effect we expected (i.e., that participants would opt for a pronoun to avoid incurring a repeated name penalty; Gordon et al., 1993). Instead, it may have had the reverse effect, priming participants to continue using names. As such, both of these forms of prominence manipulations should be included in Fossard et al.'s referential complexity scale, whereby an additional pronoun decreases referential complexity relative to an additional name (see Figure 12 for updated scale). These findings open the door for future work to implement a fully crossed manipulation of the use of an additional name versus pronoun across both same- and different-gender contexts by creating materials in which a pronoun is not ambiguous even in the same-gender context.

Our results also revealed that participants did not distinguish between the 2 → 2/same- and different-gender trials (Figure 11), unlike the findings from the pilot experiment in Arnold and

Table 8
Descriptive Statistics for Experiment 4

Participants	*M*	*SD*	Range
2 → 2 same gender			
All participants	.37	.48	0–1
Younger adults	.34	.48	0–1
Middle-aged	.49	.50	0–1
Older adults	.20	.40	0–1
2 → 2 different gender			
All participants	.40	.49	0–1
Younger adults	.35	.48	0–1
Middle-aged	.50	.50	0–1
Older adults	.26	.44	0–1
1 → 2 different gender			
All participants	.66	.47	0–1
Younger adults	.64	.48	0–1
Middle-aged	.74	.44	0–1
Older adults	.55	.50	0–1
2 → 1 different gender			
All participants	.56	.50	0–1
Younger adults	.54	.50	0–1
Middle-aged	.65	.48	0–1
Older adults	.40	.49	0–1

Note. The table shows rate of pronominal use in each condition for all participants as well as by age group (determined by a tertile age split). For the purposes of this table, we carried out a tertile age split of the data such that younger adults were between the ages of 18–31, middle-aged between 32 and 57, and older adults between 58 and 73.

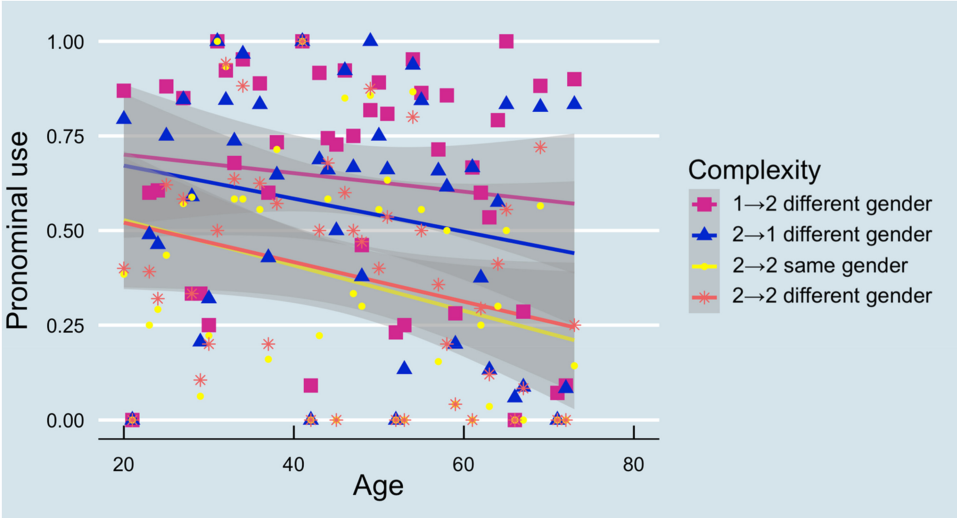
Table 9
Model Output for Pronominal Use in Experiment 4

Fixed effect	Coefficient	*SE*	*p*
Age	−.4228	.3205	.1871
1 → 2 different gender	2.5732	.4059	<.001
2 → 1 different gender	1.4157	.2563	<.001
2 → 2 same gender	−.5217	.3372	.1219
1 → 2 Different Gender × Age	.2692	.2051	.1893
2 → 1 Different Gender × Age	.0362	.1706	.8319
2 → 2 Same Gender × Age	−.1065	.2050	.6036

Note. Significant main effects and interactions are shaded. The reference level is 2 → 2/different gender.

Figure 11

Pronominal Use by Age for Each Condition in Experiment 4

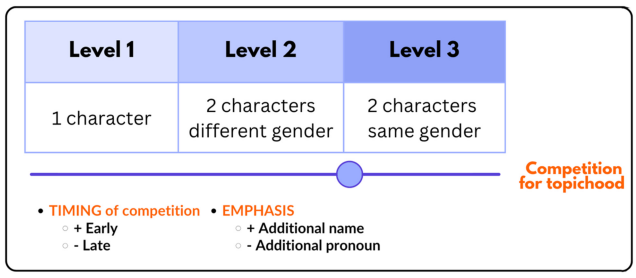


Note. Regression lines reflect the best fit of data, points reflect mean pronominal use for each age tested. The shaded bands around the regression lines represent a 95% confidence region for the regression fit. Note that prominence manipulations for 1 → 2 and 2 → 1 different-gender trials involved the use of an additional pronoun whereas the manipulations for 2 → 2 same and different-gender trials involved the use of a repeated name. See the online article for the color version of this figure.

Griffin (2007) but corresponding to more recent work (e.g., Fossard et al., 2018; Sandoz et al., 2020, 2023). This could be due to a decreased likelihood of referential ambiguity in Maintenance contexts and/or because of our emphasis on the main character's prominence, in this case through a repeated name. That is, the repeated name may have boosted the main character's prominence to such an extent that the secondary character referent was not considered a probable referent. This in turn could have resulted in speakers' making comparable prominence estimates for scenes with characters of the same or different gender.

Figure 12

Illustration of Fossard et al.'s (2018) Referential Complexity Scale Adapted to Include Our Critical Conditions



Note. Here the arrow depicts a sliding scale whereby the level of complexity of the scene/discourse can be swayed based on two factors relating to COMPETITION FOR TOPICHOOD: TIMING and EMPHASIS. If the "timing" of the competitor's presence occurs early (in Panel 1), this will move the slider up the scale, signaling an increase (+) in referential complexity; if it occurs late, there will be a decrease (−). Similarly, when emphasis is placed on the main character through a name there is an increase (+) in complexity versus when a pronoun is used, and complexity decreases (−). See the online article for the color version of this figure.

General Discussion

The longstanding assumption in the literature, based on previous work, is that older adults' referential skills deteriorate with age, likely due to a decline in executive function skills (Hendriks et al., 2008, 2014; Pratt et al., 1989). However, little research has been devoted to distinguishing which pragmatic language skills are more or less sensitive to aging—or, in other words, which linguistic abilities are preserved in old age and may or may not depend on cognitive resources. Our study directly addresses this question by testing a large sample of adults varying widely in age ($N = 496$, ages 18–82) across four experiments. Our results provide new insights into the relationship between aging, reference, and cognition, by identifying a strong link between older adults' attention-switching skills and the use of pronouns as a discourse marker of topic continuity in Maintenance contexts. Likewise, our results shed light on what type of contextual information is prioritized at different ages, highlighting older adults' preserved sensitivity to cues in the immediate visual context but a lower sensitivity to linguistic prominence cues on their own, compared to younger adults. Finally, our article contributes to an ever-growing and timely literature on referential complexity by testing and expanding upon Fossard et al.'s (2018) scale through the identification of new factors relevant to complexity.

In Experiment 1, our major finding was a steady shift to an age-related reliance on attention switching to guide pronominal use. Attention switching is a cognitive skill often overlooked in the psycholinguistics literature (in favor of inhibition and working memory accounts) but has been found to be the preferred mode of

cognitive control for older adults (e.g., Braver, 2012). Older adults' preference for this mode of control, which involves a stimulus-driven response to changes in the environment, maybe an especially suitable means of calculating the prominence of a referent, as it would involve shifting focus according to which referents are presented in the immediate visual context. That is, if there is only one referent, older adults may respond by focusing all their attention on that subject, whereas if two referents appear, they may attend to both referents, later inhibiting the secondary referent in favor of the main character as a late corrective mechanism.

Having to monitor the scene, select, and maintain focus on a particular referent (whether in the presence of competitors or not) would likely impose strong cognitive demands on older adults whose executive function skills are in decline. As such, that may explain why attention switching was strongly linked to older adults' use of pronouns (over names) as a marker of the main character's topichood. Indeed, if older adults prioritize marking topic continuity for the listener through pronominal use, then they are more likely to rely heavily on executive control to do so.

Conversely, younger adults (and to a lesser extent middle-aged adults) who pronominalized less than older adults, may prioritize using names as a default communicative heuristic based on the notion that explicit referential forms remove any potential for referential ambiguity. This type of strategy would place lower demands on executive control, which may explain why none of our cognitive measures predicted the use of pronouns in the younger and middle-aged groups. Interestingly, in Hendriks et al. (2014) younger adults also used far fewer pronouns than older adults for maintenance. The authors acknowledged that this finding was unexpected but interpreted it to reflect younger adults' motivation to make it easier for the listener to identify the intended referent. While this interpretation may be plausible in the context of their study (which involved topic shifts), our study did not involve topic shifts. Therefore, it seems more likely that in our study older adults used pronouns for their specific function as a discourse marker of topic continuity while younger adults relied on a different strategy (i.e., that of general ambiguity avoidance, by defaulting to names).

In Experiment 2, we found that the presence of a second competitor does not further increase referential complexity, at least to the extent that it would influence adults' referential choice. To test this, we manipulated the NUMBER of competitors in the scene/discourse (from 0 to 2) as a means of varying COMPETITION FOR TOPICHOOD between the main character and competitors. This resulted in one-, two-, and three-character critical trials. Here we found that both younger and older adults distinguished between scenes with one versus two characters and one versus three characters, but not between two versus three characters. In other words, pronominalization rates did not differ according to the "number" of competitors (of a different sex/gender) in the scene/discourse. As such, we conclude that this particular factor is not relevant to Fossard et al.'s referential complexity scale. In addition, from a developmental perspective, no age-related differences were found in response to these manipulations, suggesting that sensitivity to these contextual constraints remains stable over the adult lifespan.

In Experiment 3, we found that across the adult lifespan decisions regarding COMPETITION FOR TOPICHOOD between the main character and competitors are made early (i.e., in Panel 1). Indeed, by varying the TIMING of competitors' presence, we found significant differences in pronominal use such that there was a large decrease in pronominal use when there were two characters in Panel 1 as opposed to one character. As such, the "timing" of competitors' presence is highly relevant for referential complexity and should therefore be incorporated into the referential complexity scale (see Figure 12 below for updated scale). Interestingly, compared to younger adults, older adults showed a higher rate of pronominal use for trials with one character in the initial scene. This pattern may reflect a motivation to maintain attention on the target referent when there is only one character in the scene, in line with the notion that pronouns act as a discourse marker of topic continuity in Maintenance contexts (Sandoz et al., 2023).

Regarding the gender manipulation, we found that the potential for referential ambiguity did not influence pronominalization rates in Maintenance contexts. Instead, the number of characters in the initial scene played a bigger role in guiding referential choice. Nevertheless, our design was not fully crossed, therefore we addressed this issue in Experiment 4. There we again found no differences based on the sex/gender of the characters. Nevertheless, in that experiment, we also emphasized the main character's prominence through re-mention, which in turn could have resulted in speakers' making comparable prominence estimates for scenes with characters of the same or different gender. As such, while our findings tentatively suggest that participants of varying ages do not differentiate between the same- and different-gender trials in Maintenance contexts (as found in previous work, e.g., Fossard et al., 2018; Sandoz et al., 2020, 2023), follow-up work is needed to corroborate these initial results on adults' referential choices over the lifespan.

In Experiment 4, we also found that differences in pronominal rates emerged according to a decreased EMPHASIS on competitors (through re-mentioning of the main character), such that there was an increase in pronominal use for an additional pronoun versus an additional name. This is likely because an additional pronoun boosts prominence in two ways: (a) through frequency of mention and (b) pronominalization, the latter of which signals that a referent is still the topic (Arnold, 1998; Arnold et al., 2009; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). An alternative explanation is that the additional name primed participants to continue using a name, despite being suboptimal in that context (see Gordon et al., 1993 regarding the repeated name penalty). Since both explanations are plausible, each type of prominence manipulation (i.e., additional pronoun and additional name) should be included in Fossard et al.'s referential complexity scale (see Figure 12 for updated scale).

In addition, in Experiment 4, we also found age-related differences in overall pronominal use. In contrast to Experiments 1–3, there was a tendency for greater pronominal use in younger rather than older adults. We interpret this finding in light of a major difference between Experiments 1–3 and Experiment 4: in Experiment 4, the manipulations were linguistic-only and did not involve changes to the immediate visual context. It is therefore possible that this result reflects an age-related difference in sensitivity to linguistic prominence manipulations, whereby older adults are less sensitive to these contextual cues than younger adults. Together with the previous experiments, our findings can be interpreted to suggest that older adults remain sensitive to the visual cues in the immediate context but do so to a lesser extent with linguistic-only cues.

Future work should directly test this possibility by teasing apart the roles of linguistic and visual cues in adults' referential choice.

Taken together, our findings not only highlight the role of cognitive control and aging on referential choice, but also contribute to a

growing body of work that shows that referential choices are related to differences in specific stages of the discourse and at different levels of referential complexity. Expanding on Fossard et al.'s scale of referential complexity, we have added two new conditions related to COMPETITION FOR TOPICHOOD which we found to significantly impact referential choice over the adult lifespan. As shown in Figure 12, both the "timing" of competitors' presence (early or late in the scene/discourse) and the "emphasis" on competitors (relative to the main character) sway the perceived level of complexity on the scale in one direction or another and should thus be incorporated into future studies on referential complexity and referential choice. Furthermore, future work should continue to expand on the referential complexity scale by not only testing the visual properties of the scene, but also linguistically cued information, as our study shows that those properties of the discourse also play a role.

Although our study naturally has limitations, we believe this work has the potential to move the field forward. While we were unable to test adults' cognitive skills in Experiments 2–4 (since we changed from in-person to online testing due to the global pandemic), we still found interesting patterns of referential behavior in both general pronominal use and pronominal use in response to the changing referential complexity manipulations. These patterns align with our findings from Experiment 1 (i.e., age-related differences in general pronominal use but not typically in response to referential complexity). While it is possible that participants in Experiments 2–4 did not differ in cognitive control as a function of their age (as was found in Experiment 1), it is unlikely given the wide range of ages tested across three developmental stages (younger adults, middle-aged adults, and older adults). Furthermore, despite the difference in how the task was administered after Experiment 1, the materials were consistent (with slight alterations based on our experimental manipulations) and the setup in which the participant viewed the panels and was asked to imagine they were speaking to a 5-year-old child remained constant. As such, it is unlikely that major changes occurred due to the move to web-based testing. At the same time, we encourage future work to examine whether these types of tasks produce comparable results in-person and online as that could open the door for a much wider pool of participants, not only in one's own country but also globally, offering new cross-linguistic and cross-cultural data.

To conclude, in order to advance our understanding of language, cognition, and aging, it is necessary to examine referential behavior in a broad range of contexts to determine where older and younger adults' linguistic behavior diverges. Our study contributes to this research goal by offering novel results and interpretations that allow us to refine current theories of aging and referential choice, which can be applied to computational cognitive models of reference (e.g., Hendriks, 2016) and expanded into new areas of research (e.g., communicative efficiency; Long, Rohde, & Rubio-Fernández, 2020; Tal et al., 2023). In addition, our study provides a large sample of normative data which can be used to inform clinical practice. As such, this work is important for both theoretical and practical purposes, and we encourage other researchers to continue to expand and explore this line of research.

References

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics*, *99*, 78–95. https://doi.org/10.1016/j.pragma.2016.05.001

Ariel, M. (1990). *Accessing noun phrase antecedents*. Routledge.

Arnold, J. E. (1998). *Reference form and discourse patterns* [Doctoral dissertation]. Stanford University.

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, *110*(2), 131–146. https://doi.org/10.1016/j.cognition.2008.10.016

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*(4), 521–536. https://doi.org/10.1016/j.jml.2006.09.007

Bambini, V., Van Looy, L., Demiddele, K., & Schaeken, W. (2021). What is the contribution of executive functions to communicative-pragmatic skills? Insights from aging and different types of pragmatic inference. *Cognitive Processing*, *22*(3), 435–452. https://doi.org/10.1007/s10339-021-01021-w

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Chiew, K. S., & Braver, T. S. (2017). Context processing and control in the human brain: From gating models to dual mechanisms. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (1st ed., pp. 143–166). John Wiley & Sons.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Craik, F. I. M., Morris, R. G., & Gick, M. L. (1990). Adult age differences in working memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 247–267). Cambridge University Press. https://doi.org/10.1017/CBO9780511665547.014

Daniels, K., Toth, J., & Jacoby, L. (2006). The aging of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 96–111). Oxford University Press.

Fossard, M., Achim, A. M., Rousier-Vercruyssen, L., Gonzalez, S., Bureau, A., & Champagne-Lavau, M. (2018). Referential choices in a collaborative storytelling task: Discourse stages and referential complexity matter. *Frontiers in Psychology*, *9*, Article 176. https://doi.org/10.3389/fpsyg.2018.00176

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581–590. https://doi.org/10.3758/BF03192728

Fukumura, K., Hyönä, J., & Scholfield, M. (2013). Gender affects semantic competition: The effect of gender in a non-gender-marking language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1012–1021. https://doi.org/10.1037/a0031215

Fukumura, K., & Van Gompel, R. P. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, *62*(1), 52–66. https://doi.org/10.1016/j.jml.2009.09.001

Fukumura, K., Van Gompel, R. P., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *Quarterly Journal of Experimental Psychology*, *63*(9), 1700–1715. https://doi.org/10.1080/17470210903490969

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory*, *16*(7), 773–787. https://doi.org/10.1080/09658210802261124

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347. https://doi.org/10.1207/s15516709cog1703_1

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology*, *7*, Article 506. https://doi.org/10.3389/fpsyg.2016.00506

Hendriks, P., Englert, C., Wubs, E., & Hoeks, J. (2008). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information*, *17*(4), 443–466. https://doi.org/10.1007/s10849-008-9065-6

Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, *29*(4), 391–407. https://doi.org/10.1080/01690965.2013.766356

Hendriks, P., & Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, *13*(4), 319–348. https://doi.org/10.1207/s15327817la1304_3

Hindle, J. V., Martin-Forbes, P. A., Bastable, A. J., Pye, K. L., Martyr, A., Whitaker, C. J., Craik, F. I., Bialystok, E., Thomas, E. M., Mueller Gathercole, V. C., & Clare, L. (2015). Cognitive reserve in Parkinson's disease: The effects of Welsh-English bilingualism on executive function. *Parkinson's Disease*, *2015*, Article 943572. https://doi.org/10.1155/2015/943572

Jimura, K., & Braver, T. S. (2010). Age-related shifts in brain activity dynamics during task switching. *Cerebral Cortex*, *20*(6), 1420–1431. https://doi.org/10.1093/cercor/bhp206

Jorm, A. F., Windsor, T. D., Dear, K. B. G., Anstey, K. J., Christensen, H., & Rodgers, B. (2005). Age group differences in psychological distress: The role of psychosocial risk factors that vary with age. *Psychological Medicine*, *35*(9), 1253–1263. https://doi.org/10.1017/S0033291705004976

Karimi, H. (2022). Greater entropy leads to more explicit referential forms during language production. *Cognition*, *225*, Article 105093. https://doi.org/10.1016/j.cognition.2022.105093

Karmiloff-smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, *1*(1), 61–85. https://doi.org/10.1080/01690968508402071

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology*, *22*(4), 580–585. https://doi.org/10.1037/0012-1649.22.4.580

Light, L. L., Capps, J. L., Singh, A., & Albertson Owens, S. A. (1994). Comprehension and use of anaphoric devices in young and older adults. *Discourse Processes*, *18*(1), 77–103. https://doi.org/10.1080/01638539409544885

Long, M., Rohde, H., Oraa-Ali, M., & Rubio-Fernandez, P. (2023, March 3). *Referential complexity*. https://osf.io/7dzta/?view_only=04a2c177e2564b27a6699cde4ce5bbdd

Long, M., Rohde, H., & Rubio-Fernández, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports*, *10*(1), Article 8214. https://doi.org/10.1038/s41598-020-64475-6

Long, M., Vega-Mendoza, M., Rohde, H., Sorace, A., & Bak, T. H. (2020). Understudied factors contributing to variability in cognitive performance related to language learning. *Bilingualism: Language and Cognition*, *23*(4), 801–811. https://doi.org/10.1017/S1366728919000749

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition*, *170*, 25–30. https://doi.org/10.1016/j.cognition.2017.09.004

Lubben, J. E., & Gironda, M. E. (2003). Centrality of social ties to the health and well-being of older adults. In L. Berkman & L. Harooytan (Eds.), *Social work and health care in an aging world* (pp. 319–350). Springer.

Meiran, N., & Gotler, A. (2001). Modelling cognitive control in task switching and ageing. *European Journal of Cognitive Psychology*, *13*(1–2), 165–186. https://doi.org/10.1080/09541440042000269

Messer, R. H. (2015). Pragmatic language changes during normal aging: Implications for health care. *Healthy Aging & Clinical Care in the Elderly*, *7*, Article 1. https://doi.org/10.4137/HACCE.S22981

Novogrodsky, R., & Edelson, L. R. (2016). Ambiguous pronoun use in narratives of children with autism spectrum disorders. *Child Language Teaching and Therapy*, *32*(2), 241–252. https://doi.org/10.1177/0265659015602935

Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From bilingualism to bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition*, *21*(4), 867–879. https://doi.org/10.1017/S1366728918000020

Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral Cortex*, *18*(5), 1010–1028. https://doi.org/10.1093/cercor/bhm135

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Pratt, M. W., Boyes, C., Robins, S., & Manchester, J. (1989). Telling tales: Aging, working memory, and the narrative cohesion of story retellings. *Developmental Psychology*, *25*(4), 628–635. https://doi.org/10.1037/0012-1649.25.4.628

Pratt, M. W., & MacKenzie-Keating, S. (1985). Organizing stories: Effects of development and task difficulty on referential cohesion in narrative. *Developmental Psychology*, *21*(2), 350–356. https://doi.org/10.1037/0012-1649.21.2.350

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology*, *41*(4), 661–671. https://doi.org/10.1037/0012-1649.41.4.661

Renkema, J., & Schubert, C. (2018). *Introduction to discourse studies*. John Benjamins Publishing. https://doi.org/10.1075/z.219

Reuter, E.-M., Vieluf, S., Koutsandreou, F., Hübner, L., Budde, H., Godde, B., & Voelcker-Rehage, C. (2019). A non-linear relationship between selective attention and associated ERP markers across the lifespan. *Frontiers in Psychology*, *10*, Article 30. https://doi.org/10.3389/fpsyg.2019.00030

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The test of everyday attention (TEA)* (pp. 197–221). Thames Valley Test Company.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The test of everyday attention. *Journal of the International Neuropsychological Society*, *2*(6), 525–534. https://doi.org/10.1017/S1355617700001697

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, *29*(8), 912–927. https://doi.org/10.1080/01690965.2013.854918

Sandoz, M., Iglesias, K., Achim, A. M., Démonet, J. F., & Fossard, M. (2020). Referential adjustment during discourse production in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *42*(7), 710–724. https://doi.org/10.1080/13803395.2020.1798883

Sandoz, M., Iglesias, K., Achim, A. M., & Fossard, M. (2023). The contribution of discursive and cognitive factors in referential choices made by elderly people during a narrative task. *Aging, Neuropsychology, and*

Cognition. Advance online publication. https://doi.org/10.1080/13825585.2022.2150141

Scholman, M. C. J. (2019). *Coherence relations in discourse and cognition: Mapping approaches, annotations, and interpretations* [Doctoral dissertation]. University of Saarland.

Schubotz, L., Özyürek, A., & Holler, J. (2019). Age-related differences in multimodal recipient design: Younger, but not older adults, adapt speech and co-speech gestures to common ground. *Language, Cognition and Neuroscience*, *34*(2), 254–271. https://doi.org/10.1080/23273798.2018.1527377

Serratrice, L. (2013). The role of number of referents and animacy in children's use of pronouns. *Journal of Pragmatics*, *56*, 31–42. https://doi.org/10.1016/j.pragma.2013.05.008

Shukla, V., Long, M., & Rubio-Fernandez, P. (2022). Children's acquisition of new/given markers in English, Hindi, Mandinka and Spanish: Exploring the effect of optionality during grammaticalization. *Glossa Psycholinguistics*, *1*(1), 1–42. https://doi.org/10.5070/G6011120

Strawson, P. F. (1964). Identifying reference and truth values. *Theoria*, *30*(2), 96–118. https://doi.org/10.1111/j.1755-2567.1964.tb00404.x

Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant references with language learners: Evidence for communicatively-efficient referential choice. *Journal of Memory and Language*, *128*, Article 104378. https://doi.org/10.1016/j.jml.2022.104378

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*, 40–46. https://doi.org/10.1016/j.cognition.2014.12.008

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience & Biobehavioral Reviews*, *26*(7), 849–857. https://doi.org/10.1016/s0149-7634(02)00071-4

Veríssimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., & Ullman, M. T. (2022). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nature Human Behaviour*, *6*(1), 97–110. https://doi.org/10.1038/s41562-021-01169-7

Von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. https://doi.org/10.1080/01690965.2012.728232

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: A meta-analysis. *Psychology and Aging*, *26*(1), 15–20. https://doi.org/10.1037/a0020912

Waters, G., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 531–555). Lawrence Erlbaum Associates.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 51–79. https://doi.org/10.1080/713755607

Wright, H. H. (Ed.). (2016). *Cognition, language and aging*. John Benjamins Publishing.

(*Appendices follow*)

Appendix A

Initial Cognitive Assessment to Determine Age-Related Variability Within Our Sample.

Participants

A total of 209 adults were recruited from the University of Edinburgh volunteer panel, the University Careers Services website, and local community groups. All correspondence was done via email, which ensured that everyone who registered was computer literate. Written informed consent was obtained from all participants and the study was approved by the University of Edinburgh Linguistics and English Language ethics committee. Prior to analysis, we excluded nine participants: six nonnative English speakers, one participant with abnormally low attentional scores, one participant with hearing difficulty, and one due to a technical malfunction. Thus, we report results from 200 native English speakers ages 19–82 with normal vision and hearing (see Figure A1 for the age distribution within the sample).

Cognitive Tasks

Test of Everyday Attention

Participants' attentional skills were measured via the Test of Everyday Attention (TEA; Robertson et al., 1994), a clinical test based on Posner and Petersen's (1990) multisystem attentional model. The test offers a fine-grained method of assessing an individual's cognitive resources by separating attention into theoretically distinct factors through the use of three auditory subtests: attention switching, inhibition, and sustained attention. The test is sensitive enough to detect subtle differences in attention and has been standardized through a normative sample of healthy adults ages 18–80 (Robertson et al., 1996). The normative data demonstrate that participants' performance varies greatly across the scoring range (Robertson et al., 1996), with the exception of the sustained attention task (the least computationally difficult of the three measures) in which adults generally make few errors (Robertson et al., 1994). Participants who respond incorrectly on two or more items for the sustained attention measure (equivalent to a score of 71% or lower), are considered to have abnormally low levels of attention (Robertson et al., 1994). More recently this test has been used across a variety of linguistic studies, from language learning (Vega-Mendoza et al., 2015) and bilingualism (Hindle et al., 2015; Ooi et al., 2018) to perspective-taking (Long et al., 2018) and referential communication (Long, Rohde, & Rubio-Fernández, 2020).

For each of the measures (referred to as "the elevator tasks"), participants are asked to envision that they have entered an elevator on the ground floor. The floor light indicator does not work, so in order to know which floor they are on they must count the tones they hear. After each trial, they are asked which floor they are on. Performance on each subtest is measured as the percentage of trials with correct responses (0–100).

Elevator Task (Sustained Attention, $n = 7$ Trials). Participants are presented with tones of the same pitch at irregular intervals and must keep track of the count. As healthy individuals are expected to perform near ceiling, this task was used as a baseline measure.

Elevator Task With Distraction (Selective Attention/Inhibition, $n = 10$ Trials). Participants are presented with low and high tones (mimicking the experience of being on an elevator and hearing a beep as it moves from floor to floor). However, rather than counting all of the tones (to assess which floor they are on), participants must selectively attend to and count the low tones only while ignoring interspersed high tones. Performing well requires inhibiting high tones while counting low tones or selectively attending to low tones only.

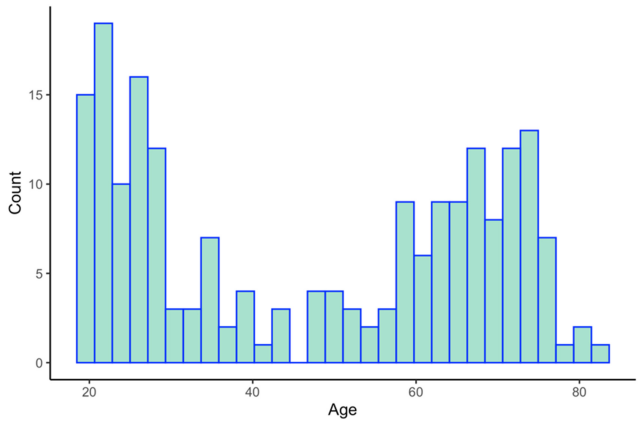
Elevator Task With Reversal (Attention Switching, $n = 10$ Trials). Participants are presented with low, medium, and high tones in random order (again mimicking the experience of being on an elevator and hearing a beep as it moves floor to floor). In order to know which floor they are on, they must count medium tones only (e.g., upon hearing one medium beep they will move from the ground floor to the second floor). Low tones indicate the elevator will change direction and move down with the subsequent medium tones (thus can be thought of as a down arrow, signally the direction of movement with the following medium tones), while high tones indicate the elevator will move up with subsequent medium tones (thus can be thought of as an up arrow). Performing well requires inhibiting low and high tones from the count (as they indicate direction and not the floor you are on) while efficiently disengaging inhibition and refocusing attention upon hearing a middle tone.

Following standard practice, we calculated the percentage of trials with correct responses (0–100) for each elevator task. As participants receive one score for each task, we were unable to calculate a by-trial split-half reliability for these measures. However, previous work has demonstrated a high test–retest reliability after 1 week (Robertson et al., 1996), which suggests that these measures are stable over time.

Reading Span

Verbal working memory (VWM) capacity was assessed using an automated reading span task (Scholman, 2019) adapted from Waters

Figure A1
Histogram of Age Within the Sample



Note. See the online article for the color version of this figure.

(*Appendices continue*)

et al. (1987) and Waters and Caplan (1996). Participants were presented with a total of 56 sentences varying in length from eight to 11 words and were asked to judge the acceptability of each sentence by clicking "acceptable" or "unacceptable" on a computer screen (see bottom of Appendix A for a full list of sentences). Four different sentence types were presented, with varying levels of complexity to see how well people could remember the content. Half of the sentences contained a verb that required an animate subject, while the other half contained a verb that required an animate object. Furthermore, half of the sentences were grammatically acceptable and the other half were made unacceptable by reversing the animacy of the subject and object noun phrases (e.g., *It was the boy that the hat bought*).

After a set of two to five sentences (recommended by Conway et al., 2005), participants were asked to recall the last word of each sentence in the order in which it was presented by typing the words on the computer. Sets were presented in random order so participants could not anticipate how many words they would be asked to recall. Before beginning the task, participants were given a practice phase in which they judged the acceptability of eight sentences. Reaction times were measured for each of the practice trials. A participant's average reaction time during the practice trials plus two standard deviations was used as a time-out during the task to ensure participants went at their own pace but were not taking extra time to rehearse the words (cf., Unsworth et al., 2005; von der Malsburg & Vasishth, 2013). Participants were also given a practice phase for the recall portion of the task. In the practice phase, participants judged sentences and recalled the last words for a block of three sentences and a block of five sentences. Practice trials were not used in the analysis.

Working memory was calculated using the partial-credit unit scoring procedure where the partial-credit unit represents the mean proportion of words within a set that were recalled correctly (Conway et al., 2005; Friedman & Miyake, 2005). Typos were accounted for by allowing a one-character difference between the target word and response. Previous work has demonstrated a high split-half and test–retest reliability for this measure (Conway et al., 2005).

Statistical Analysis

The α level for all reported tests was set to $p < .05$ and all analyses were run using R statistical software (R Core Team, 2019). As expected, participants performed at ceiling on the baseline measure of sustained attention (mean score = 99.21; after the exclusion of one participant, score = 57), thus this will not be considered further. To assess the relationship between participants' age and cognitive abilities, we ran linear regression models in which we modeled age as a predictor for each of the following cognitive skills: switching, VWM, and inhibition.

Results

Descriptive statistics are reported in Table A1. Note that while some participants scored 0 on the inhibition and switching tasks, all participants passed the baseline measure of sustained attention. Furthermore, a certain level of task difficulty is necessary in order to capture variability in participants' performance (and avoid a homogenous ceiling effect). As demonstrated in Table A1, the scores for inhibition and switching range from 0 to 100, providing a full spectrum of variability across participants.

Table A1
Performance on the Cognitive Tasks for all Participants as Well as a Median Age Split for Younger and Older Adults

Participants	*M*	*SD*	Range
Sustained attention			
All participants	99.21	3.26	85.71–100
Younger adults	98.88	3.84	85.71–100
Older adults	99.56	2.14	85.71–100
Inhibition			
All participants	83	22.83	0–100
Younger adults	80.59	24.41	0–100
Older adults	85.51	20.76	10–100
Switching			
All participants	58.25	35.39	0–100
Younger adults	69.31	30.41	0–100
Older adults	46.73	36.54	0–100
Verbal working memory			
All participants	.74	.17	.06–1
Younger adults	.78	.14	.32–1
Older adults	.69	.18	.06–.97

Note. For the purposes of this table, we carried out a median age split by dividing our sample ($n = 200$) by the median age (49) such that younger adults were between the ages of 19–49 and older adults were between the ages of 50–82. The purpose of this was to provide descriptive statistics regarding age and cognitive functions. In the regression models, however, age was treated as a continuous variable.

Our results revealed a main effect of age on switching ($\beta = -0.608$, $SE = .112$, $t = -5.431$, $p < .001$) and VWM ($\beta = -0.002$, $SE = 0.0005$, $t = -4.022$, $p < .001$) in the predicted direction: as age increased, switching and VWM decreased (see Figure A1). These results provide confidence that there is sufficient age-related variability within our sample. Moreover, the direction of the effect is in line with findings from the aging literature (Gilchrist et al., 2008; Wasylyshyn et al., 2011). While our measure of inhibition showed no effect of age ($\beta = .100$, $SE = .077$, $t = 1.298$, $p = .196$), this result supports findings from a recent meta-analysis calling into question the notion that inhibition typically deteriorates with age (Rey-Mermet & Gade, 2018). Another possibility for this null effect of age is that our inhibition measure was not sensitive enough. However, as shown in Table A1 and Figure A2, scores were not uniformly at ceiling (as was found for the baseline measure of sustained attention), thus our participants displayed variability in this measure.

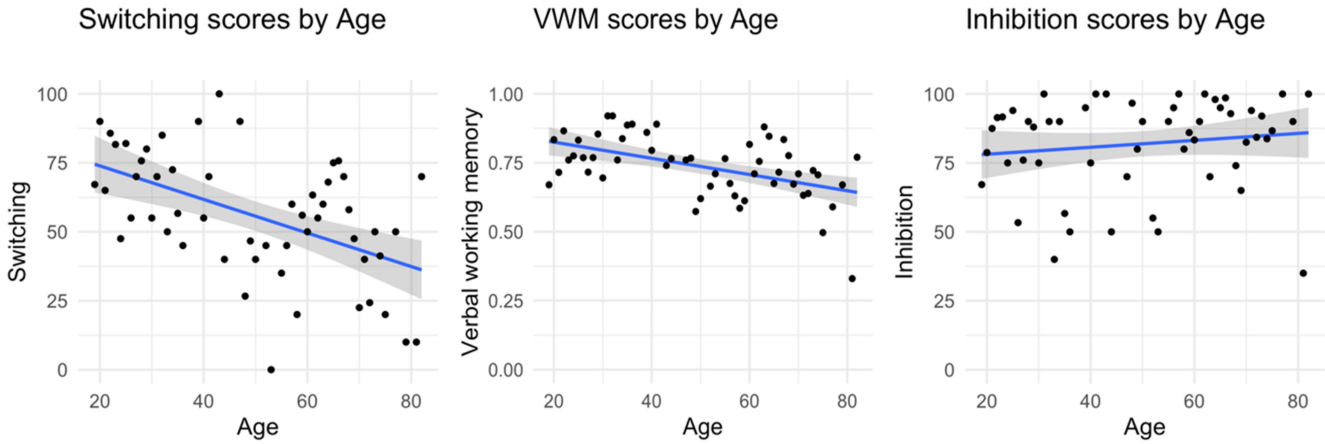
Full list of experimental items from the reading span task

1. It was the elephant that escaped from the zoo. (Animate subj., cleft subj., acceptable.)
2. It was the professor that forgot the handout. (Animate subj., cleft subj., acceptable.)
3. It was the man that clenched the pillow. (Animate subj., cleft subj., acceptable.)
4. It was the bear that made the growling sound. (Animate subj., cleft subj., acceptable.)
5. It was the sound that made the computer. (Animate subj., cleft subj., unacceptable.)
6. It was the document that filed the librarian. (Animate subj., cleft subj., unacceptable.)

(*Appendices continue*)

Figure A2

Switching, Verbal Working Memory, and Inhibition Scores by Age



Note. Regression lines reflect the best fit of data, points reflect the mean score for each age tested. The shaded bands around the regression lines represent a 95% confidence region for the regression fit. See the online article for the color version of this figure.

7. It was the cookie that ate the talented dancer. (Animate subj., cleft subj., unacceptable.)
8. It was the toy that fascinated the child. (Animate obj., cleft subj., acceptable.)
9. It was the porcelain doll that scared the little girl. (Animate obj., cleft subj., acceptable.)
10. It was the bike that hit the news reporter. (Animate obj., cleft subj., acceptable.)
11. It was the building that impressed the architect. (Animate obj., cleft subj., acceptable.)
12. It was the sulky teenager that bored the book. (Animate obj., cleft subj., unacceptable.)
13. It was the florist that astonished the bouquet. (Animate obj., cleft subj., unacceptable.)
14. It was the family that shocked the revelation. (Animate obj., cleft subj., unacceptable.)
15. It was the teddy bear that the child wanted. (Animate subj., cleft obj., acceptable.)
16. It was the Polish bread that the family liked to eat. (Animate subj., cleft obj., acceptable.)
17. It was the upbeat pop song that the grandfather hated. (Animate subj., cleft obj., acceptable.)
18. It was the young man that the engagement ring bought. (Animate subj., cleft obj., unacceptable.)
19. It was the CEO that the enormous file requested. (Animate subj., cleft obj., unacceptable.)
20. It was the owner of the house that the light fixed. (Animate subj., cleft obj., unacceptable.)
21. It was the model that the cheeseburger ate. (Animate subj., cleft obj., unacceptable.)
22. It was the older man that the speech offended. (Animate obj., cleft obj., acceptable.)
23. It was the widow that the vivid dream tormented. (Animate obj., cleft obj., acceptable.)
24. It was the innocent people that the restrictions affected. (Animate obj., cleft obj., acceptable.)
25. It was the berry that the amateur hunter poisoned. (Animate obj., cleft obj., unacceptable.)
26. It was the memory that the heartbroken girl haunted. (Animate obj., cleft obj., unacceptable.)
27. It was the embroidery that the young girl calmed down. (Animate obj., cleft obj., unacceptable.)
28. It was the amulet that the innocent children hypnotized. (Animate obj., cleft obj., unacceptable.)
29. The boy envied the friend that bought a new game. (Animate subj., obj.-subj., acceptable.)
30. The girl played with the friend that injured her foot. (Animate subj., obj.-subj., acceptable.)
31. The athlete hired the manager that applied for the position. (Animate subj., obj.-subj., acceptable.)
32. The scarf loved the designer that kept the model warm. (Animate subj., obj.-subj., unacceptable.)
33. The milk drank the lawyer that turned sour. (Animate subj., obj.-subj., unacceptable.)
34. The castle hated the man that interested all tourists. (Animate subj., obj.-subj., unacceptable.)
35. The sound listened to the man that came from the basement. (Animate subj., obj.-subj., unacceptable.)
36. The artwork fascinated the girl that loved to paint. (Animate obj., obj.-subj., acceptable.)
37. The drug worried the pharmacist that worked with patients. (Animate obj., obj.-subj., acceptable.)
38. The book inspired the young girl that admired strong women. (Animate obj., obj.-subj., acceptable.)
39. The flower field charmed the lady that had always lived in the city. (Animate obj., obj.-subj., acceptable.)
40. The woman bothered the rain that had straightened her hair. (Animate obj., obj.-subj., unacceptable.)
41. The police officer puzzled the evidence that investigated the case. (Animate obj., obj., unacceptable.)
42. The gardener pleased the good weather that worked outside. (Animate obj., obj.-subj., unacceptable.)
43. The employee that the boss fired yelled at her supervisor. (Animate subj., subj.-obj., acceptable.)
44. The singer that everybody loved hated rock music. (Animate subj., subj.-obj., acceptable.)

(*Appendices continue*)

45. The actress that the audience booed at called her manager. (Animate subj., subj.-obj., acceptable.)
46. The man that the painter loved despised his parents. (Animate subj., subj.-obj., acceptable.)
47. The door that nobody trusted pushed the doctor open. (Animate subj., subj.-obj., unacceptable.)
48. The secrets that the woman was seeing betrayed her psychologist. (Animate subj., subj.-obj., unacceptable.)
49. The knee that the coach trained injured his gymnast. (Animate subj., subj.-obj., unacceptable.)
50. The meat that the butcher cut delighted the customer. (Animate obj., subj.-obj., acceptable.)
51. The pen that the teacher brought splashed ink on the children. (Animate obj., subj.-obj., acceptable.)
52. The song that the mother played comforted the sad daughter. (Animate obj., subj.-obj., acceptable.)
53. The man that the doctor recommended helped the exercises. (Animate obj., subj.-obj., unacceptable.)
54. The man that the plumber didn't fix annoyed the leaking shower. (Animate obj., subj.-obj., unacceptable.)
55. The customers that the woman made impressed the bracelet. (Animate obj., subj.-obj., unacceptable.)
56. The audience that the acrobats performed astonished the trick. (Animate obj., subj.-obj., unacceptable.)

Appendix B

Table B1
Model Output for Pronominal Use and Cognitive Control in Experiment 1

Fixed effect	Coefficient	*SE*	*p*
VWM	−.187843	.38696	.85699
Switching	.79506	.32927	.01575
Inhibition	−.32213	.29086	.26808
Number of characters	−3.63489	.34722	<.0001
Age	.94772	.31862	.00293
Version	−.44732	.62783	.47616
VWM × Number of Characters	.34981	.25099	.16340
Switching × Number of Characters	−.26665	.25211	.29021
Inhibition × Number of Characters	.14920	.25931	.56502
VWM × Age	−.60016	.30346	.47962
Switching × Age	−.18122	.34036	.59422
Inhibition × Age	−.07961	.29564	.78770
Age × Number of Characters	−.18163	.25539	.47697
VWM × Version	−1.34090	.60560	.26820
Switching × Version	−.26053	.65567	.69111
Inhibition × Version	.32923	.58129	.57114
Number of Characters × Version	−1.12074	.62288	.07198
Age × Version	−1.57740	.63537	.13044
VWM × Age × Number of Characters	−.18448	.24402	.44964
Switching × Age × Number of Characters	.01720	.26204	.94767
Inhibition × Age × Number of Characters	−.18398	.24474	.45221
VWM × Number of Characters × Version	.60055	.50365	.23311
Switching × Number of Characters × Version	−.07647	.50629	.87994
Inhibition × Number of Characters × Version	−.75460	.49952	.13088
VWM × Age × Version	.57926	.60465	.33806
Switching × Age × Version	.106744	.68133	.11718
Inhibition × Age × Version	−.52146	.58127	.36966
Number of Characters × Age × Version	.14299	.53170	.78798
VWM × Number of Characters × Age × Version	−.63599	.49048	.19474
Switching × Number of Characters × Age × Version	−.34106	.53724	.52488
Inhibition × Number of Characters × Age × Version	−.36555	.50270	.46712

Note. VWM = verbal working memory.

(*Appendices continue*)

Appendix C

Full list of experimental items from the story continuation tasks in Experiments 1–4:

Critical trials from Experiment 1:

1. Doggie cooked rice (with Mousey) for dinner.
2. Mousey was out all day (with Birdie) and came home hungry.
3. Froggy decided to work on a painting (with Duckie) at the weekend.
4. Bunny wrote a song (with Horsey) to play on the piano.
5. Birdie ran out the door (with Kitty) and forgot to turn off the bath.
6. Duckie walked to the botanical gardens (with Piggy) to look at the plants.
7. Horsey was feeling ill when walking (with Goosey) to work.
8. Kitty went (with Froggy) to buy a motorbike.
9. Piggy went for a run (with Bunny) around the park.
10. Goosey sat down (with Doggie) on the park bench.
11. Doggie went (with Mousey) to pick up the newspaper.
12. Duckie took a cooking course (with Piggy) over the summer.
13. Froggy made a cake (with Duckie) to celebrate the occasion.
14. Goosey went ice skating (with Doggie) in the city.
15. Birdie ran the marathon (with Kitty) in New York.
16. Kitty went (with Froggy) to buy a new camera.
17. Horsey went to the countryside (with Goosey) for a peaceful afternoon.
18. Mousey went to the canal (with Birdie) during lunch break.
19. Piggy went (with Bunny) to the outdoor concert.
20. Bunny went (with Horsey) to the café.

Filler trials from Experiment 1:

1. Kitty just finished a quilting project.
2. Birdie spent the evening with Bunny at the dance school.
3. Goosey wanted to try something new.
4. Bunny decided to invite a friend out to lunch.
5. Doggie played hide and seek with Mousey this afternoon.
6. Horsey went with Bunny to buy some flowers.
7. Kitty went apple-picking with Piggy in the orchard.
8. Froggy had a sore ankle after running on the cobblestone.
9. Kitty needed help with the homework.
10. Doggie learned how to tie a shoelace with Mousey at school.
11. Horsey went with Bunny to the chemist.
12. Mousey was excited to go to the theme park.
13. Bunny went with Birdie to the art show.
14. Froggy wanted to join a community project.
15. Kitty did some Spring cleaning with Doggie at the weekend.
16. Piggy needed help packing for a holiday abroad.
17. Horsey watched the sunset with Bunny on the terrace.
18. Goosey went bowling with Horsey after school.
19. Duckie realized that the shirt had wrinkles.
20. Goosey went with Birdie to band practice.
21. Bunny got lost when going for a walk.
22. Duckie went hiking with Froggy and tripped on a rock.
23. Goosey made dinner with Horsey this evening.
24. Froggy decided to get in shape with Bunny over the summer.
25. Bunny worked on the puzzle with Piggy for 10 days straight.
26. Birdie decided to chop some firewood.
27. Froggy decided to go for a long walk since the weather was nice.
28. Goosey thought that everyone forgot about the birthday party.
29. Froggy went with Kitty to the lake this morning.
30. Duckie studied for the exam with Birdie yesterday.
31. Duckie heard that an amazing band was in town.
32. Piggy got ready with Mousey for the celebration.
33. Duckie was very upset about the decision.
34. Kitty went to play in the back garden.
35. Duckie went with Froggy to the park to throw the frisbee.
36. Mousey spent hours playing a board game with Birdie last night.
37. Kitty decided to go outside and enjoy the day.
38. Horsey had a craving for something sweet.
39. Doggie tried to think of something interesting to do after school.
40. Mousey forgot to make the bed this morning.

Critical trials from Experiment 2:

1. Goosey attended the awards night (with Froggy) (and also Horsey).
2. Birdie forgot to turn off the bath as he ran out the door (with Kitty) (and also Mousey).
3. Duckie walked to the botanical gardens to look at the plants (with Piggy) (and also Horsey).
4. Bunny went to the café (with Horsey) (and also Piggy).
5. Froggy made a cake to celebrate the occasion (with Duckie) (and also Goosey).
6. Kitty went to buy a new camera (with Froggy) (and also Birdie).
7. Kitty decided to go camping at the weekend (with Doggie) (and also Piggy).
8. Mousey went to the canal during lunch break (with Birdie) (and also Doggie).
9. Doggie was feeling cold while sitting in the living room (with Duckie) (and also Bunny).
10. Piggy went to the outdoor concert (with Bunny) (and also Duckie).
11. Duckie took a cooking course over the summer (with Piggy) (and also Birdie).
12. Horsey went to the countryside for a peaceful afternoon (with Goosey) (and also Kitty).
13. Mousey was hungry after a long walk (with Birdie) (and also Piggy).

(*Appendices continue*)

14. Horsey was feeling ill when walking to work (with Goosey) (and also Bunny).
15. Piggy went to wash a load of laundry (with Kitty) (and also Bunny).
16. Birdie ran the marathon in New York (with Kitty) (and also Mousey).
17. Kitty went to buy a motorbike (with Froggy) (and also Doggie).
18. Doggie went to pick up the newspaper (with Mousey) (and also Duckie).
19. Froggy decided to work on a painting at the weekend (with Duckie) (and also Bunny).
20. Goosey went to the park for some fresh air (with Doggie) (and also Froggy).
21. Mousey went to the pool on a hot summer day (with Froggy) (and also Doggie).
22. Doggie cooked rice for dinner (with Mousey) (and also Kitty).
23. Piggy went for a run around the park (with Bunny) (and also Kitty).
24. Goosey went ice skating in the city (with Doggie) (and also Froggy).
25. Bunny decided to test out a new song (with Horsey) (and also Doggie).
26. Duckie saw a shooting star (with Birdie) (and also Piggy).
27. Kitty went to the shop to buy stationery (with Horsey) (and also Birdie).
28. Horsey went to the supermarket to do his weekly shopping (with Bunny) (and also Goosey).
29. Froggy tried out for the talent show (with Duckie) (and also Mousey).
30. Birdie had a notion for black tea and walked into the kitchen (with Mousey) (and also Goosey).

1 → 2 different-gender trials from Experiment 3:

1. Doggie tried to think of something interesting to do after school.
2. Kitty just finished a quilting project.
3. Bunny decided to invite a friend out to lunch.
4. Goosey wanted to try something new.
5. Mousey was excited to go to the theme park.
6. Froggy wanted to join a community project.
7. Piggy needed help packing for a holiday abroad.
8. Duckie realized that the shirt had wrinkles.
9. Birdie decided to chop some firewood.
10. Horsey had a craving for something sweet.

2 → 1 different-gender trials from Experiment 3:

1. Birdie spent the evening with Bunny at the dance school.
2. Doggie played hide and seek with Mousey this afternoon.
3. Horsey went with Bunny to buy some flowers.
4. Kitty went apple-picking with Piggy in the orchard.
5. Bunny went with Birdie to the art show.
6. Goosey went bowling with Horsey after school.
7. Duckie went hiking with Froggy and tripped on a rock.
8. Froggy decided to get in shape with Bunny over the summer.
9. Piggy got ready with Mousey for the celebration.
10. Mousey spent hours playing a board game with Birdie last night.

2 → 2 same-gender trials from Experiment 3:

1. Bunny went with Goosey to the cafe.
2. Froggy made a cake with Horsey to celebrate the occasion.
3. Doggie went with Piggy to pick up the newspaper.
4. Goosey went ice skating with Kitty in the city.
5. Horsey went to the countryside with Birdie for a peaceful afternoon.
6. Mousey went to the canal with Bunny during lunch break.
7. Piggy went with Froggy to the outdoor concert.
8. Duckie took a cooking course with Mousey over the summer.
9. Birdie ran the marathon with Doggie in New York.
10. Kitty went with Duckie to buy a new camera.

1 → 2 different-gender trials from Experiment 4:

1. Doggie tried to think of something interesting to do after school. He had an idea.
2. Kitty just finished a quilting project. She was pleased with the patchwork.
3. Bunny decided to invite a friend out to lunch. She went outside to make a call.
4. Goosey wanted to try something new. She thought of an activity that seemed interesting.
5. Mousey was excited to go to the theme park. She likes all of the attractions.
6. Froggy wanted to join a community project. He liked to help out and make new friends.
7. Piggy needed help packing for a holiday abroad. He wasn't sure what clothes to bring to Italy.
8. Duckie realized that the shirt had wrinkles. She needed to sort it out before leaving for work.
9. Birdie decided to chop some firewood. He wanted some logs for the fireplace.
10. Froggy had a sore ankle after running on the cobblestone. He thought it had been fractured.
11. Kitty needed help with the homework. She found the questions quite difficult.
12. Bunny got lost when going for a walk. She wasn't sure how to get back home.
13. Froggy decided to go for a long walk since the weather was nice. He ventured to the other side of town.
14. Goosey thought that everyone had forgotten about the birthday party. She had spent ages planning it and was very sad.
15. Duckie heard that an amazing band was in town. She decided to buy tickets.
16. Duckie was very upset about the decision. She couldn't believe it.
17. Kitty went to play in the back garden. She was in a great mood.
18. Kitty decided to go outside and enjoy the day. She came up with a fun plan.
19. Mousey forgot to make the bed this morning. She needed some help with it.

(*Appendices continue*)

20. Horsey had a craving for something sweet. He has quite the sweet tooth.

2 → 1 different-gender trials from Experiment 4:

1. Birdie spent the evening with Bunny at the dance school. He loved to go to dance class.
2. Doggie played hide and seek with Mousey this afternoon. He thought of the perfect spot to hide in.
3. Horsey went with Bunny to buy some flowers. He wanted something big for the living room.
4. Kitty went apple-picking with Piggy in the orchard. She loved to find the best ones.
5. Bunny went with Birdie to the art show. She was hoping to find a nice painting.
6. Goosey went bowling with Horsey after school. She had been practicing for weeks.
7. Duckie went hiking with Froggy and tripped on a rock. She was quite clumsy.
8. Froggy decided to get in shape with Bunny over the summer. He even bought a gym membership.
9. Piggy got ready with Mousey for the celebration. He wasn't sure which tie to wear.
10. Mousey spent hours playing a board game with Birdie last night. She really wanted to win so wouldn't stop playing.
11. Doggie learned how to tie a shoelace with Mousey at school. He was so excited to be able to do it.
12. Horsey went with Bunny to the chemist. He had a bad headache and needed ibuprofen.
13. Kitty did some spring cleaning with Doggie at the weekend. She loved to tidy up.
14. Horsey watched the sunset with Bunny on the terrace. He was amazed by the vivid colors.
15. Goosey went with Birdie to band practice. She had been practicing all week.
16. Goosey made dinner with Horsey this evening. She wanted to prepare a salad.
17. Bunny worked on the puzzle with Piggy for 10 days straight. She was determined to finish it.
18. Froggy went with Kitty to the lake this morning. He decided a boat ride would be nice.
19. Duckie studied for the exam with Birdie yesterday. She wasn't feeling very confident.
20. Duckie went with Froggy to the park to throw the frisbee. She was hoping to get better at the game.

2 → 2 same-gender trials from Experiment 4:

1. Bunny went with Goosey to the cafe. The cafe was Bunny's favorite place.
2. Froggy made a cake with Horsey to celebrate the occasion. Making cakes is Froggy's forte.
3. Doggie went with Piggy to pick up the newspaper. There was always something interesting for Doggie to read.
4. Goosey went ice skating with Kitty in the city. Ice skating was Goosey's most beloved pastime.
5. Horsey went to the countryside with Birdie for a peaceful afternoon. Getting out in the fresh air relaxed Horsey.
6. Mousey went to the canal with Bunny during lunch break. Going there was Mousey's idea.
7. Piggy went with Froggy to the outdoor concert. The concert exceeded Piggy's expectations.
8. Duckie took a cooking course with Mousey over the summer. Learning to cook was one of Duckie's main priorities.
9. Birdie ran the marathon with Doggie in New York. The marathon was difficult for Birdie.
10. Kitty went with Duckie to buy a new camera. The new camera was much more sophisticated than Kitty's old one.

2 → 2 same-gender trials from Experiment 4:

1. Doggie cooked rice with Mousey for dinner. Rice was Doggie's favorite meal.
2. Goosey sat down with Doggie on the park bench. Work had been quite exhausting for Goosey that day.
3. Horsey was feeling ill when walking with Goosey to work. Something was aggravating Horsey's allergies.
4. Mousey was out all day with Birdie and came home hungry. A growling sound came from Mousey's stomach.
5. Froggy decided to work on a painting with Duckie at the weekend. Painting was Froggy's new hobby.
6. Duckie walked to the botanical gardens with Piggy to look at the plants. The daisies drew Duckie's attention.
7. Kitty went with Froggy to buy a motorbike. Riding a bike had been Kitty's childhood dream.
8. Piggy went for a run with Bunny around the park. It had been months since Piggy's last run.
9. Bunny wrote a song with Horsey to play on the piano. The song had a special meaning for Bunny.
10. Birdie ran out the door with Kitty and forgot to turn off the bath. It was time for Birdie's favorite TV show.

Received March 7, 2021
Revision received May 23, 2023
Accepted May 26, 2023 ■