JSLHR

Research Article

Effects of Aging on Interference During Pronoun Resolution

Jana Reifegerste[a] and Claudia Felser[a]

Purpose: The purpose of this study is to investigate the effects of healthy aging on the ability to suppress grammatically illicit antecedents during pronoun resolution.
Method: In 2 reading-based acceptability–judgment experiments, younger and older speakers of German read sentences containing an object pronoun and 2 potential antecedent noun phrases, only 1 of which was a grammatically licit antecedent. Using a gender-mismatch paradigm, we compared to what extent younger and older speakers were sensitive to feature (mis)matches between the pronoun and either of the 2 antecedents. All participants were fluent readers of German and had finished at least secondary education.
Results: Experiment 1 used a self-paced reading paradigm. Older speakers showed greater sensitivity than younger ones to mismatching licit antecedents, but no group showed any evidence of interference from an intervening competitor antecedent. In Experiment 2, we increased the processing demand by using paced word-by-word stimulus presentation and longer sentences. Here, older participants showed reduced sensitivity, in comparison with younger people, to mismatching licit antecedents. Unlike our younger participants, they showed signs of distraction by the presence of a linearly closer but grammatically inappropriate antecedent when no appropriate antecedent was available.
Conclusion: Together, our results show that older (but not younger) speakers' ability to compute intrasentential referential dependencies is vulnerable to increased task demands. We briefly discuss a potential role for executive functions, such as interference control.

Aging has been shown to affect language processes, which draw on general cognitive abilities subject to age-related decline. This leads to performance decline in tasks tapping lexical access, such as picture naming, naming from definitions, or lexical decisions (Allen, Madden, & Crozier, 1991; Bowles & Poon, 1981, 1985; Connor, Spiro, Obler, & Albert, 2004; Feyereisen, 1997; Newman & German, 2005), for example. Regarding sentence-level processing, the picture is less clear. Some studies reported evidence for preserved syntactic processing (Davis, Zhuang, Wright, & Tyler, 2014; Tyler et al., 2010), whereas others observed age effects for complex grammatical phenomena, such as ambiguity resolution in garden-path sentences (Kemtes & Kemper, 1997) or the processing of complex embeddings (Kemper, 1986). Recent findings by Reifegerste, Hauer, and Felser (2017) indicate that older speakers' decreased processing abilities can compromise their ability to establish morphosyntactic dependencies. In this study, we investigate how aging affects the ability to compute intrasentential referential dependencies by examining the resolution of German object pronouns.

Pronouns are referentially dependent elements, which can be fully interpreted only when linked to an appropriate antecedent or discourse referent. Successful pronoun resolution requires the integration of structural, semantic, and discourse-level information. Encountering a pronoun during text or discourse comprehension is thought to trigger a memory search for an appropriate antecedent or discourse referent (see, e.g., Cunnings, Patterson, & Felser, 2015). According to cue-based memory-access models of comprehension (Lewis & Vasishth, 2005; McElree, 2000; McElree, Foraker, & Dyer, 2003), the antecedent search is feature-driven or cue-driven. All potential antecedents within the current mental sentence or discourse representation may be evaluated in parallel, with the one that matches the pronoun best in terms of, for example, its gender and number features ultimately being selected.

The resolution of object pronouns, such as *her* in (1) is also sensitive to configurational cues (e.g., Chow, Lewis, & Phillips, 2014). In the linguistic literature, configurational cues that restrict the interpretation of pronominals

[a]Potsdam Research Institute for Multilingualism, University of Potsdam, Brandenburg, Germany
Correspondence to Jana Reifegerste: jana.reifegerste@uni-potsdam.de
Editor-in-Chief: Sean Redmond
Editor: Carl Coelho
Received May 18, 2017
Revision received August 8, 2017
Accepted August 23, 2017
https://doi.org/10.1044/2017_JSLHR-L-17-0183

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

are traditionally captured by Chomsky's (1981) binding theory. Consider the sentence in (1) below for illustration.

(1) *Suzy thinks that Mary dislikes her.*

The pronoun *her* in (1) can refer to either the matrix subject *Suzy* or a semantically or pragmatically suitable referent outside the sentence. Importantly, however, a phrase-structure, sensitive constraint known as *binding Condition B* (Chomsky, 1981) rules out the local subject *Mary* as a possible antecedent, even though it matches the pronoun's gender and number features. That is, the sentence in example (1) is incompatible with a reader's inference that Mary dislikes herself. From the processing perspective, applying Condition B requires that comprehenders ignore or dismiss grammatically illicit competitor antecedents, such as *Mary* in (1).

Findings from younger adults show that Condition B is usually respected during real-time sentence reading and auditory sentence comprehension, suggesting that structurally inappropriate antecedents are not considered (Clifton, Kennison, & Albrecht, 1997; Nicol & Swinney, 1989; Patterson, Trompelt, & Felser, 2014). Some studies, however, have reported interference effects, with the processing of object pronouns being adversely affected by the presence of a feature-matching illicit antecedent, such as *Mary* in (1) (Badecker & Straub, 2002; Kennison, 2003). Similarity-based interference effects indicate that Condition B can be overridden during the initial antecedent search or might come into play only at later processing stages. The relative salience of a linearly closer competitor antecedent (*Mary* in Example 1) may be boosted by a decay over time of the more distant antecedent's (such as *Suzy*) memory representation (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006).

Previous work on pronoun resolution in less commonly studied populations, such as children (Avrutin & Wexler, 1992; Chien & Wexler, 1990; Clackson, Felser, & Clahsen, 2011; Guasti, 2002; Koster, 1993; Sigurjónsdóttir & Hyams, 1992), second-language speakers (Kim, Montrul, & Yoon, 2015; Patterson et al., 2014), and individuals with aphasia (Grodzinsky, Wexler, Chien, Marakovitz, & Solomon, 1993; Ruigendijk, Vasić, & Avrutin, 2006), indicates that interpreting pronouns may incur considerable processing cost, suggesting that it may become more difficult with increasing age. Considering that pronoun resolution requires keeping track of discourse referents and storing representations of potential antecedents in memory, it is conceivable that older people, whose memory and cognitive control abilities may be reduced (Dobbs & Rule, 1989; Hartshorne & Germine, 2015; Hoogendam, Hofman, van der Geest, van der Lugt, & Ikram, 2014; Manard, Carabin, Jaspar, & Collette, 2014; Verhaeghen & Salthouse, 1997), find it more difficult than younger ones to identify a suitable antecedent or to ignore an illicit competitor antecedent. Light and Anderson (1985) asked participants to read paragraphs and respond to questions referring to a pronoun in the final sentence. Older speakers made significantly more errors in this task than younger speakers. Studies on the resolution of ambiguous pronouns found that older speakers are able to use contextual information; however, they may make more mistakes than younger speakers, particularly if the sentence providing the context and the pronoun are separated by other sentential material (Leonard, Waters, & Caplan, 1997; Light & Capps, 1986). We are not aware of any published studies examining the application of Condition B in aging.

This Study

We conducted two acceptability–judgment experiments in which younger and older participants judged sentences containing an object pronoun and two potential antecedents. In Experiment 1, we used self-paced reading to establish whether participants apply Condition B when reading sentences containing object pronouns at their own pace. Assessing the possible role of individual differences in working memory (WM) on pronoun resolution was a secondary goal. In Experiment 2, we used externally paced serial visual presentation to assess the role of increased processing demands during pronoun resolution. Keeping the gender of the pronoun constant[1], we manipulated the gender features of both the licit and a linearly closer but grammatically illicit antecedent, resulting in four experimental conditions (see Table 1).

All experimental sentences contained a male or a female name in matrix subject position (noun phrase 1 [NP1]) and a finite declarative complement clause introduced by the complementiser *dass* "that." The embedded subject (noun phrase 2 [NP2]) contained kinship terms or professions for which German has distinct lexical variants for each gender. This means that unlike in previous studies, which relied on stereotypical gender (mis)matches (e.g., Cunnings & Felser, 2013; Shake & Stine-Morrow, 2011), in this study, both the grammatical and the definitional gender of NP2 was always unambiguously masculine or feminine. NP1 was always a grammatically licit antecedent, whereas NP2 was always grammatically illicit.

Sentences in which NP1's gender features mismatch the pronoun's gender (c, d) should be judged as incorrect more often than those containing a matching NP1 (a, b) because, in the former case, the pronoun lacks a licit gender-matched sentence-internal antecedent. Effects of NP1-gender—in the absence of any NP2-gender effects—would suggest that Condition B prevents illicit antecedents from being considered. Effects of NP2-gender, in contrast, would indicate that grammatically inappropriate antecedents are considered, a prediction in line with cue-based direct-access models of comprehension.

We compared the performance of a younger student sample and a group of older people. Regarding age effects, three possible outcomes seem plausible. First, if the ability to compute structurally mediated referential dependencies is unaffected by aging, younger and older people should show the same judgment rates and response times (RTs) across

[1]All sentences contained a masculine singular object pronoun because the German singular feminine pronoun *sie* is homonymous with the gender-neutral plural pronoun.

Table 1. Overview of the four experimental conditions, Experiment 1.

Condition	Example	
a. Double match	*Erik verkündet, dass der Opa ihm geschrieben hat.*	Erik announces that the grandpa him written has
b. NP1 match, NP2 mismatch	*Erik verkündet, dass die Oma ihm geschrieben hat.*	Erik announces that the grandma him written has
c. NP1 mismatch, NP2 match	*Jasmin verkündet, dass der Opa ihm geschrieben hat.*	Jasmin announces that the grandpa him written has
d. Double mismatch	*Jasmin verkündet, dass die Oma ihm geschrieben hat.*	Jasmin announces that the grandma him written has
	"Erik/Jasmin announces that grandpa/grandma has written to him."	

Note. NP1 = noun phrase 1; NP2 = noun phrase 2.

experimental conditions. Second, we may find main effects of age across conditions, especially in participants' RTs, reflecting older people's generally slower processing speed. Last, we might find specific effects of age, that is, differences in the acceptance rates and/or in RT patterns between younger and older people, reflecting between-group differences in the ability (a) to notice a mismatching licit antecedent and/or (b) to suppress the activation of grammatically illicit competitor antecedents. This would be indicated by interactions involving the factor age.

Experiment 1

Method

Participants

We recruited 30 younger and 30 older community-dwelling native speakers of German. All participants gave written informed consent, reported (corrected-to-)normal vision and hearing, and no neurological/language-related impairments. No participant had learned another language before the age of 11 years. All participants were living in Germany at the time of testing and had not spent more than a year in a non-German-speaking country. Although more women than men participated, the gender ratio did not differ between the younger and the older group ($\chi^2 = 0.635$, *ns*). Educational level of the two groups ranged from 10 to 17 years (secondary education to master's degree). See Table 2 for demographic information.

In order to assess participants' WM skills, we administered a German version of the reading span task developed by Waters and Caplan (1996). Participants were asked to rate 40 (semantically plausible or implausible) sentences. The sentences were presented in 11 blocks, with between two and six sentences per block. Participants read the sentence and then provided a plausibility judgment. After each block, they were asked to write down the last word of every sentence in the correct order of appearance. Participants scored 0.5 points for every correct plausibility judgment and 0.5 points per correctly recalled word, up to a total of 40 points.

Table 2. Demographic information about the participants in Experiment 1.

Category	Younger group			Older group		
Gender	17 women, 13 men			20 women, 10 men		
Handedness	29 right, 1 left			28 right, 2 left		
	M	*SD*	Range	*M*	*SD*	Range
Age	25.6	4.7	20–40	68.3	5.9	60–80
Education (in years)	13.6	1.8	12–17	14.3	2.8	10–17
Working memory score	35.8	2.7	27.5–39.5	32.5	3.1	27.0–38.5

Materials

We constructed 48 experimental sentences as in Table 1. See Appendix A for the materials. In order to ensure that all participants considered the first names used as NP1s as unambiguously male or female, participants filled out a questionnaire that asked them to rate 96 names on a scale of 1 to 5 (1 = *unambiguously male,* 3 = *equally male or female,* 5 = *unambiguously female*) after the experiment. The 48 names used in this study were rated as unambiguously male or female by all participants.

Ninety-six filler sentences of similar average length and varying grammatical complexity were added, half of which were grammatical and semantically plausible, whereas the other half was ungrammatical. The ungrammatical filler sentences contained grammatical violations, such as wrong case (e.g., a noun in dative case following a preposition requiring accusative case), mismatch in adjective–noun case agreement, or overregularizations of irregular verb forms. Many of the acceptable sentences contained a female name as NP1, so as to avoid biasing participants toward responding "no" to sentences containing a female subject.

Apparatus and Procedure

The sentences were presented word-by-word using the noncumulative moving-window paradigm (Just, Carpenter, & Woolley, 1982). For stimuli presentation and data recording, we used Linger (http://tedlab.mit.edu/~dr/Linger/), with participants controlling the presentation duration of each word via button press. After the presentation of the last word of each sentence, three question marks appeared on the screen, prompting the participant to decide whether the sentence was acceptable, with the dominant hand controlling the "yes" button. There was no timeout.

The dependent measures were acceptance rate, judgment RTs, and word-by-word reading latencies. We calculated linear mixed-effects models for inverse-transformed

RTs and generalized linear mixed-effects models (binomial family) for the acceptance rates, using the languageR package (Baayen, 2013) and the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). These two statistical packages are particularly suitable for the analysis of psycholinguistic data because they implement linear mixed-effects models using the programming environment R (R Core Team, 2013). (Generalized) linear mixed-effects models allow for the analysis of categorical (e.g., acceptability) and continuous (e.g., RT) dependent variables on the basis of independent fixed factors, while also taking into account the influence of both random by-participant and random by-item variance, providing a distinct advantage over traditionally used analyses of variance. Manipulated fixed factors were NP1-gender, NP2-gender, and age group, with participant and item as random factors. In order to control for the influence of differences in frequency and length in letters between masculine (M_{Length} = 7.2, M_{Freq} = 4.1) and feminine NP2s (M_{Length} = 9.0, M_{Freq} = 1.9), we further added NP2-frequency and NP2-length as fixed factors[2] as well as trial number (position of trial within the experiment). Because NP2-frequency, NP2-length, and trial number were not manipulated, effects involving these control variables will be reported in the tables in Appendix B but will not be discussed. All fixed factors were centered.

Results

We excluded trials with RTs shorter or longer than 2.5 *SD*s from the mean on a per-participant basis, resulting in 1.9% data loss. This cutoff criterion, which is considered standard practice in psycholinguistic experiments with healthy adults, ensures that unusually slow (e.g., due to distraction) or unusually fast responses (e.g., due to erroneous button presses) do not distort the data.

Acceptability

Figure 1 illustrates acceptability rates for end-of-sentence judgments.

Sentences containing a gender-matching matrix subject (NP1) elicited higher acceptability rates than sentences containing a mismatching one. We found an interaction between NP1-gender and age group. Although both groups considered sentences with a matrix subject whose gender matched that of the pronoun as more acceptable, the effect of NP1-gender was weaker for the younger group than the older group. There were no further main effects or interactions; see Table B1 in Appendix B for the entire model.

Judgment Latencies

For the analysis of both judgment latencies and word-by-word reading times, we excluded all trials in which participants accepted unacceptable sentences (without a matching licit antecedent) or rejected acceptable sentences (with a matching licit antecedent).[3] Figure 2 illustrates RTs for end-of-sentence judgments.

Older participants had significantly longer RTs compared to younger people. An interaction between NP1-gender and group reflected the fact that younger speakers showed longer RTs when the sentence started with a female subject compared with a male subject (t = −2.40), whereas older speakers' RTs were not affected by this manipulation (t = 0.95).

There were no further main effects or interactions; see Table B2 in Appendix B for the entire model.

Reading Times

Figure 3 illustrates the two participant groups' word-by-word reading profiles across our four experimental conditions. See Table B3 in Appendix B for an overview of the significant effects.

For all words, we found a main effect of age, with older people showing longer reading latencies than younger people. There were no further significant main effects of or interactions with manipulated variables on the first five words.

On the pronoun (Segment 6), we found a main effect of NP1-gender, indicating that readers took longer to read the pronoun when the sentence started with a gender-mismatching name than when it started with a matching one. Although the effect was significant for both groups, it was larger for older participants (231 ms difference, t = −6.88) compared with younger participants (56 ms difference, t = −3.16), as indicated by an interaction between NP1-gender and age group. We did not find any main effects of or interactions with NP2-gender (t < 1.00).

On the seventh word (pronoun + 1), we found a main effect of NP1-gender and an interaction between NP1-gender and age group. Older people showed longer latencies when NP1 was male compared with when it was female (t = 5.51); that is, the NP1-gender effect we saw at the pronoun reversed at the following word for older people. Younger people did not show an effect of NP1-gender (t = −0.89) here.

There were no further main effects or interactions. We did not analyze reading times for the final word because these are affected by wrap-up processes and may include part of the decision component.

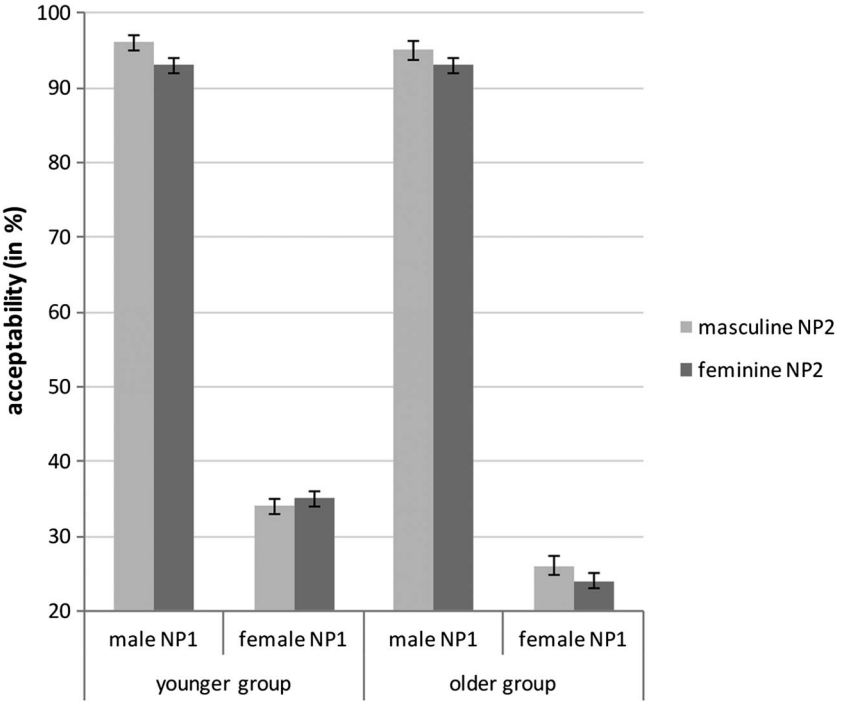
Discussion

Using the self-paced reading paradigm, Experiment 1 examined whether younger and older speakers of German were equally sensitive to (mis)matching licit and illicit antecedents for object pronouns. The results showed age

[2]We used log-transformed (natural log) frequencies based on the Mannheim corpus as reported in the Celex database (Baayen, Piepenbrock, & Gulikers, 1995).

[3]While sentences lacking a sentence-internal antecedent for a pronoun are not ungrammatical, we expected them to be judged as unacceptable. Where an unexpected "yes" response was provided, we cannot be sure whether the participant did indeed process the gender of both NP1 and the pronoun, which is essential for the task at hand. No participant accepted more than half of our unacceptable sentences, suggesting that they did not generally consider sentences without a sentence-internal antecedent acceptable.

Figure 1. Acceptability rate as a function of NP1-gender, NP2-gender, and age group in Experiment 1. In all graphs, error bars reflect the standard error of the mean. NP1 = noun phrase 1; NP2 = noun phrase 2.



effects on acceptability judgments rates, judgment RTs, and word-by-word reading latencies.

For acceptability rates, we found both groups to be more accepting of sentences with a gender-matching licit antecedent (NP1) for the pronoun than of sentences without, and no effects of, the illicit antecedent's (NP2) gender. This pattern of results is consistent with the application of Condition B, which rules out coreference between an object

Figure 2. Judgment reaction times as a function of NP1-gender, NP2-gender, and age group in Experiment 1. NP1 = noun phrase 1; NP2 = noun phrase 2.

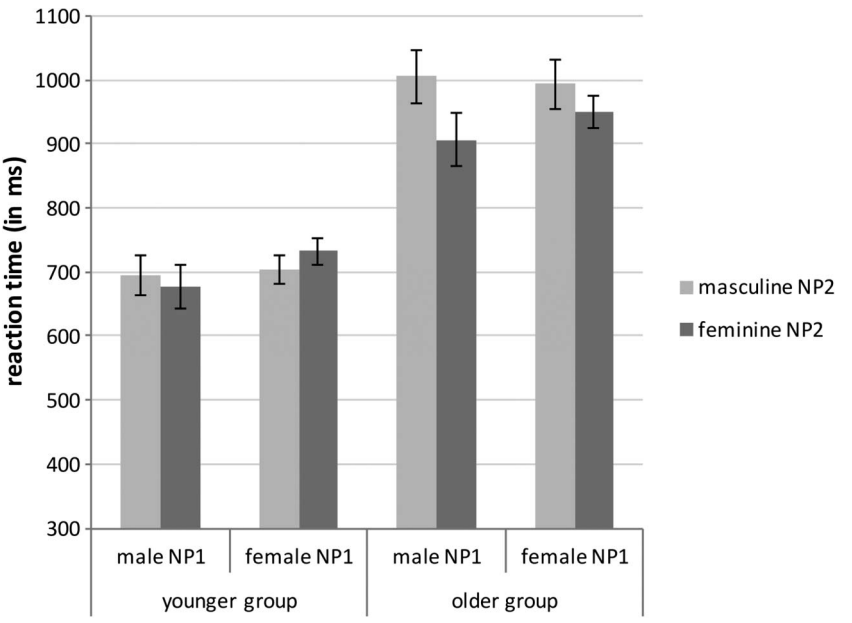
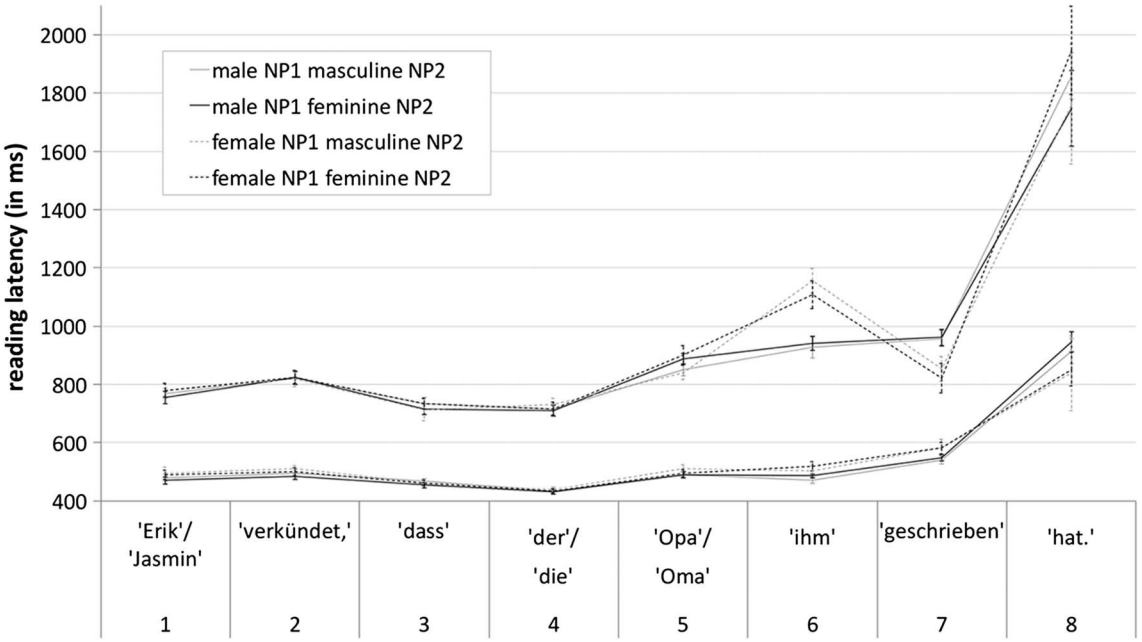


Figure 3. Word-by-word reading latencies for the two age groups (top: older group, bottom: younger group) in Experiment 1. NP1 = noun phrase 1; NP2 = noun phrase 2.



pronoun and its local subject. This effect was stronger for older speakers, who were more likely than younger speakers to reject sentences in which the pronoun lacked a feature-matching, grammatically licit antecedent. There are two possible explanations for this between-group difference: One is that younger speakers were more likely than older ones to consider the possibility of the pronoun referring to an (unmentioned) sentence-external discourse referent. A further possibility is that older speakers' longer reading times enabled them to encode a fully specified memory representation of NP1.

Participants' end-of-sentence judgment RTs showed a main effect of age, which was expected due to age-related general slowing (Verhaeghen & Salthouse, 1997). In addition, an interaction between age and NP1-gender indicated that younger, but not older speakers, took longer to reject sentences with mismatching NP1s than to accept those with matching NP1s. We believe that this difference is due to differences in reading or task strategy, such that younger speakers simply read through the sentences and evaluated their acceptability only upon encountering the end-of-trial question, whereas older speakers kept monitoring the acceptability of the stimulus sentences as they read each word (see below).

The analysis of participants' word-by-word reading times showed effects of NP1-gender on reading times at and following the pronoun. At the pronoun, we found longer reading times for sentences with gender-mismatching NP1s compared with sentences in which the matrix subject's gender matched that of the pronoun. This effect was larger for older speakers, who took an average of 199 ms longer to process a pronoun without a legitimate sentence-internal antecedent, compared with a difference of only 31 ms in the younger group. This age-related difference resembles an analogous effect reported by Reifegerste et al. (2017), who found that when reading sentences containing subject–verb agreement violations at their own pace, older participants' reading times were more strongly affected by ungrammaticality than younger speakers. In this study, this effect reversed for older speakers at the next word, where they showed shorter reading times for unacceptable sentences, that is, sentences with a female NP1. This reversal suggests that the pronoun marks the point in time at which older speakers decide whether the sentence is acceptable: When encountering a pronoun that lacked a gender-matching grammatically appropriate antecedent, they considered the sentence unacceptable and, so, did not spend a lot of time reading the remaining words of the sentence. Younger speakers, however, did not show a reversed NP1-gender effect, which—in conjunction with their longer judgment RTs for NP1-mismatch sentences—suggests that they waited until the end of a trial before making their decision.

Crucially, in none of our measures did we find effects of NP2-gender, indicating that both groups were able to block interference from a linearly closer but grammatically illicit competitor antecedent. Participants' WM score did not affect our dependent measures.

Having established that older speakers are able to apply Condition B in a self-paced reading task, we asked whether effects of aging on speakers' sensitivity to the illicit antecedent and/or on their susceptibility to interference from an illicit antecedent might have been masked by their

relatively longer reading times. Experiment 2 addresses this by using paced serial visual presentation with a timeout. In this paradigm, participants see the sentences appear word by word on the screen for a limited amount of time before making an acceptability judgment at the end of each trial. Processing load was further increased by presenting sentences that are three words longer than the sentences used in Experiment 1. Otherwise, Experiments 1 and 2 used the same design and materials so that potential differences in the findings can be attributed to differences in task demands between the experiments. Note that because WM did not appear to affect the results in Experiment 1, we did not include a WM measure in Experiment 2.

Experiment 2

Method

Participants

We recruited 30 younger and 30 older community-dwelling monolingual native speakers of German. All participants were living in Germany at the time of testing and had not spent more than a year in a non-German-speaking country. Participants gave written informed consent and reported (corrected-to-)normal vision and hearing and no neurological/language-related impairments. The gender ratio did not differ between the younger and the older groups ($\chi^2 = 2.783$, *ns*). Educational level ranged between 10 and 20 years (secondary education to doctoral degree). See Table 3 for demographic information.

Materials

The 48 experimental items were adapted from the sentences used in Experiment 1. We expanded the sentences by three words by adding an adjective preceding NP2 and a prepositional phrase modifier containing the (gender-unmarked) name of a city or country following it (e.g., *Erik verkündet, dass der nette Opa aus München ihm geschrieben hat.* "Erik announces that the nice grandpa from Munich has written to him"). These three additional words were identical across conditions.

Forty-eight filler sentences of similar average length and varying grammatical complexity were added, half of which were grammatical and semantically plausible, whereas the other half was ungrammatical.

Table 3. Demographic information about the participants in Experiment 2.

Category	Younger group			Older group		
Gender	27 women, 3 men			22 women, 8 men		
Handedness	29 right, 1 left			29 right, 1 left		
	M	*SD*	Range	*M*	*SD*	Range
Age	21.2	2.5	18–28	57.8	8.5	50–81
Education (in years)	13.3	2.0	12–17	14.7	2.6	10–20

Apparatus and Procedure

The experiment was presented on a computer using DMDX version 5.1 (Forster & Forster, 2003), a Windows-based experiment program, which allows for the presentation of stimuli and measurement of RTs with millisecond accuracy. The sentences were presented word by word in paced serial visual presentation with words appearing in the center of the screen. Note that the presentation rate of 750 ms lies considerably below the average per-word reading time older participants showed in Experiment 1. After the presentation of the last word of each sentence, three question marks appeared on the screen, prompting the participant to decide whether the sentence was acceptable. If the participant did not respond within 5 s, the next sentence appeared. Responses were recorded by button press, with the dominant hand controlling the "yes" button.

The dependent measures were acceptability rates and judgment RTs. We calculated (generalized) linear mixed-effects models in the same manner as in Experiment 1. Fixed factors were age group, NP1-gender, NP2-gender, NP2-length, NP2-frequency, and trial number. Participants and items were random factors.

Results

We excluded trials with RTs shorter or longer than 2.5 *SD*s from the mean on a per-participant basis, resulting in 2.8% data loss.

Acceptability

Figure 4 illustrates acceptability rates for end-of-sentence judgments.

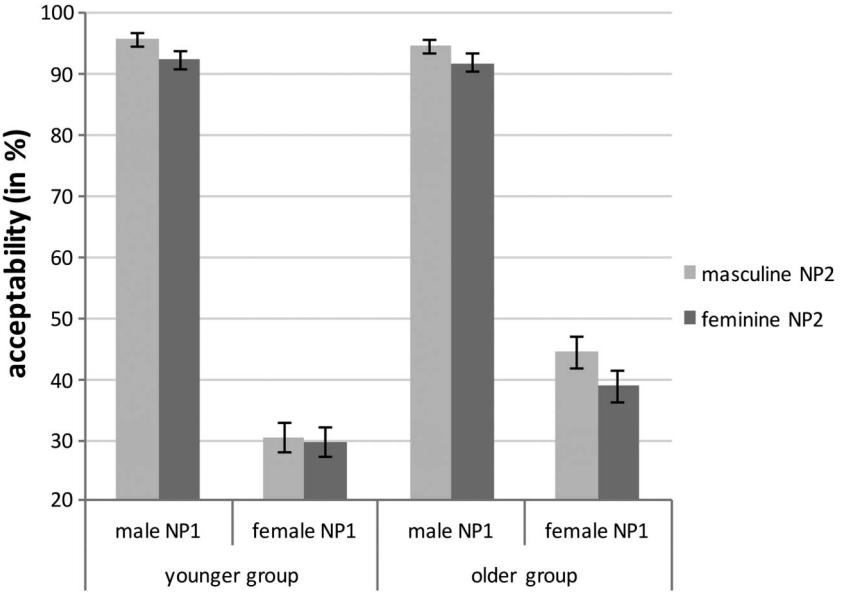
As in Experiment 1, sentences starting with a male name were rated more acceptable than sentences starting with a female name. An interaction between NP1-gender and age group indicates that this effect was stronger for the younger group than the older group. Although there was a numerical trend for masculine NP2s to increase acceptability, there were no statistical effects of or interactions with NP2-gender. See Table B4 in Appendix B for the entire model.

Response Latencies

As in Experiment 1, for the analyses of RTs, we excluded all trials in which participants accepted unacceptable sentences (without a matching licit antecedent) or rejected acceptable sentences (with a matching licit antecedent). Figure 5 illustrates RTs for end-of-sentence judgments.

As in Experiment 1, older participants showed longer RTs than younger people. Several interactions involving age indicate different RT patterns for the two groups. An interaction between NP2-gender and age indicates that while older speakers were sensitive to the gender of the embedded NP ($t = 2.93$; longer RTs for masculine NP2s), younger speakers' RTs were not affected by this manipulation ($t = 1.28$). We also found an interaction between NP1-gender and NP2-gender, with masculine NP2s leading to longer RTs (compared with feminine NP2s) only when NP1 was

Figure 4. Acceptability rate as a function of NP1-gender, NP2-gender, and age group in Experiment 2. NP1 = noun phrase 1; NP2 = noun phrase 2.

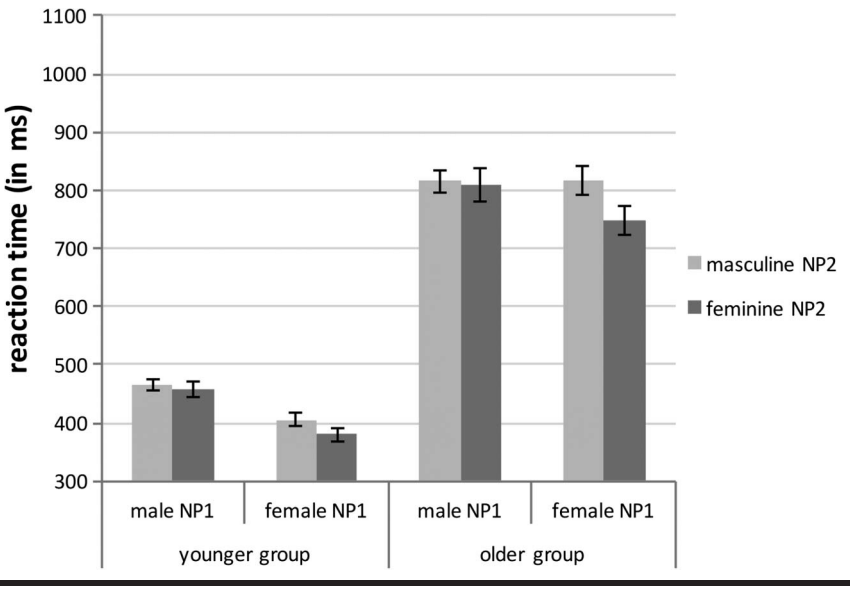


female but not when it was male. In other words, NP2-gender affected RTs only if it was impossible to link the pronoun to NP1 (the grammatically licit antecedent) but not otherwise. It is important to note that this effect was modulated by age group, as indicated by a significant three-way interaction between NP1-gender, NP2-gender, and age group. Further per-group analyses revealed that the interaction between NP1-gender and NP2-gender was significant only for the older group ($t = -3.17$) but not for the younger group ($t = -0.79$). See Table B5 in Appendix B for the entire model.

Discussion

In Experiment 2, we used the externally paced serial visual presentation paradigm to investigate whether younger and older speakers were equally sensitive to our gender manipulations in a more demanding task compared with

Figure 5. Reaction times as a function of NP1-gender, NP2-gender, and age group in Experiment 2. NP1 = noun phrase 1; NP2 = noun phrase 2.



Experiment 1. We increased the processing load in Experiment 2 by making the sentences three words longer and by limiting the amount of time participants could take to process the individual words.

The acceptability data showed evidence of sensitivity to the presence of a gender-mismatching matrix subject (NP1), whose structural position renders it a possible antecedent for the pronoun in the embedded clause, in both groups. Unlike in Experiment 1, this effect was weaker for older speakers, who were less likely to reject sentences with mismatching NP1s than younger speakers. It is conceivable that the increased processing demand in Experiment 2 (including the introduction of a timeout) led older speakers to be less "critical" of unacceptable sentences. As in Experiment 1, NP2-gender had no effect on acceptability rates in either group.

Our RT analysis revealed that older speakers were slower to respond than younger ones, which is unsurprising given that age-related slowing is well-attested (Verhaeghen & Salthouse, 1997). We also observed a more specific age effect in participants' end-of-trial judgment times, with older—but not younger—speakers showing longer RTs when the illicit antecedent (NP2) was masculine compared with when it was feminine, for sentences that lacked a gender-matching licit antecedent. This difference in RTs as a function of NP2-gender is unlikely to be the consequence of a general age-related decrease in processing speed, because such general slowing (e.g., of perceptual or motor processing) should affect processing across all conditions. Instead, we interpret this as an inhibitory interference effect: Where the pronoun could not be successfully linked to the matrix subject (NP1) because of a gender mismatch, older speakers briefly hesitated in the presence of a gender-matching but illicit competitor antecedent (NP2) before providing a negative judgment. Thus, it seems that older speakers are less efficient than younger speakers at blocking intervening competitor antecedents from interfering with the resolution of object pronouns.

General Discussion

Aging is associated with a plethora of changes to cognition. The extent to which aging affects grammatical processing seems to depend, at least in part, on the complexity of the grammatical phenomenon under study, with age effects emerging for more complex phenomena, such as subject–verb agreement computation across intervening sentence material (Reifegerste et al., 2017) or ambiguity resolution (Kemtes & Kemper, 1997). In this article, we asked whether aging affects the ability to compute sentence-internal referential dependencies and to suppress interference from a linearly closer competitor antecedent. Using two different experimental paradigms, we asked participants to judge the acceptability of sentences that contained an object pronoun preceded by both a grammatically licit and a grammatically inappropriate competitor antecedent.

In Experiment 1, we found older speakers to be less accepting than younger ones of sentences containing gender-mismatching legitimate antecedents. Younger speakers, on the other hand, took more time to reject sentences with mismatching legitimate antecedents than to accept sentences with matching antecedents, whereas this was not the case for the older speakers. This pattern suggests different strategies for younger and older speakers when reading our stimulus sentences and deciding on their acceptability, though ultimately, both groups of speakers considered sentences containing a pronoun without a gender-matching grammatically licit antecedent to be less acceptable. However, no participant group showed evidence of interference from an illicit competitor antecedent.

Experiment 2, in contrast, found older speakers to be more accepting of sentences that lacked a gender-matching legitimate antecedent for a pronoun than younger ones. It is important to note that we furthermore observed age effects on participants' end-of-trial judgment RTs, with only older speakers being slowed down in the presence of a gender-matching but grammatically inappropriate competitor antecedent. Together, these age-related differences indicate that, in Experiment 2, older speakers were less successful than younger ones in encoding or retaining a faithful representation of the matrix subject in memory and were also more likely than younger speakers to temporarily consider a grammatically unlicensed antecedent when there was no gender-matching, licit sentence-internal antecedent available.

Whereas our younger participants' performance patterns were largely consistent between the two experimental tasks and in accordance with Condition B, older speakers' performance patterns differed considerably between Experiments 1 and 2. Taken together, the results from our older participants indicate that their ability to evaluate the suitability of potential antecedents for a pronoun is affected by task demands. In Experiment 2, the paced word-by-word presentation created greater processing demand compared with the self-paced reading task used in Experiment 1, in which participants could take as much time as they wanted to process each word. The 750-ms per-word presentation time used in Experiment 2 was shorter than older speakers' average per-word reading time in Experiment 1, which we propose created processing demands high enough to uncover the temporary consideration of grammatically inappropriate licensors. Our second experiment further increased the processing load by presenting participants with sentences that were three words longer than those used in Experiment 1. Previous research found evidence that older speakers experience greater difficulty than younger speakers processing longer sentences compared with shorter sentences (Kemper, 1986). Our findings are in line with earlier findings suggesting that older speakers' grammatical processing skills are preserved when the phenomenon under study—or the task itself—puts relatively few processing demands on the aging cognitive system, compared with more cognitively demanding tasks, for which age differences may emerge (Shafto & Tyler, 2014).

The results from our older participants in Experiment 2 are problematic for the claim that Condition B precludes

grammatically inappropriate antecedents from being considered (Nicol & Swinney, 1989). The data from our older participants indicate that a syntactically unlicensed antecedent may be considered if processing demands are high. In our study, interference was selective, however, in that older speakers would consider a grammatically unlicensed antecedent only if no gender-matching legitimate antecedent was available. Our results resemble those of Chow et al. (2014), who describe sensitivity to illicit antecedents as the consequence of a repair mechanism in the absence of a suitable gender-matching licit antecedent. In our Experiment 2, when older participants encountered a pronoun that lacked a matching licit antecedent, they might have temporarily considered the illicit one in an attempt to "rescue" an otherwise unacceptable sentence. The extra processing time required for this was reflected in older speakers' end-of-trial judgment times in Experiment 2, albeit not in Experiment 1, where processing demand was comparatively reduced.

Limitations

While this study is the first to examine the application of Condition B in aging, it comes with a few limitations. First, we would like to address the finding that participants' WM score did not affect any of the dependent measures in Experiment 1. WM is regarded as one of the main causes of older speakers' decline in several grammatical processes (Grossman et al., 2002; Waters & Caplan, 2001)—in fact, the very test used here modulated age effects on agreement–attraction errors in one of our previous studies (Reifegerste et al., 2017). Considering the nature of the task employed in our study, which requires speakers to disregard a linearly closer competitor antecedent, it is possible that age-related differences in inhibition and interference control (Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999) are at the heart of the age differences found in this study (see Novick, Trueswell, & Thompson-Schill, 2005, for a review on the role of cognitive control in syntactic processing). Future studies on this topic should include a measure of inhibitory or interference control (e.g., Stroop, Corsi Blocks, and digit span) to explore the involvement of cognitive control in the processing of pronoun resolution further.

Another limitation is that our study treated age as a dichotomous factor distinguishing younger and older speakers. Although this follows previous studies on aging and language, most of which have not examined developmental trajectories at "middle age," future studies should include people between the ages of 40 and 50 years. Such a more balanced design may even allow for the treatment of age as a continuous factor rather than a categorical factor.

Clinical Implications and Future Directions

Studying language processing in healthy aging contributes importantly to the advancement of knowledge beyond the scope of basic research. Examining how healthy older people fare with the processing of syntactically mediated referential dependencies may help shed further light on syntactic processing in clinical populations, for which healthy age-matched individuals usually serve as the baseline. Difficulties with pronoun resolution have been found, for instance, for people with agrammatic Broca's aphasia (Grodzinsky et al., 1993; Ruigendijk et al., 2006). Although we are not aware of any work that has studied Condition B in aphasia in particular, our findings suggest that these individuals, who have severe processing deficits, may have problems with this constraint as well.

Moreover, we would like to note that our study investigated only sentence-internal pronoun resolution, focusing on a single structure-sensitive constraint. Pronoun resolution in discourse is guided by a much richer set of constraints, including discourse pragmatic and semantic coherence cues, as well as world knowledge (see, e.g., Kehler, Kertz, Rohde, & Elman, 2008; Rohde & Kehler, 2014; Ueno & Kehler, 2010). Future studies may want to address how (age-related) changes to processing capacities may affect these kinds of constraints on pronoun resolution.

Acknowledgments

This work was supported by an Alexander-von-Humboldt-Professorship awarded to Harald Clahsen. The authors declare no competing financial interests. The authors are grateful to the people who participated in the experiments, and thank Franziska Hauer for assistance with the data collection.

References

Allen, P. A., Madden, D. J., & Crozier, L. C. (1991). Adult age differences in letter-level and word-level processing. *Psychology and Aging, 6,* 261–271. https://doi.org/10.1037//0882-7974.6.2.261

Avrutin, S., & Wexler, K. (1992). Development of principle B in Russian: Coindexation at LF and coreference. *Language Acquisition, 2,* 259–306.

Baayen, R. H. (2013). languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics." R package version 1.4.1. Retrieved from https://CRAN.R-project.org/package=languageR

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database. Linguistic data consortium*. Philadelphia, PA: University of Pennsylvania.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphora. *Journal of Experimental Psychology: Learning Memory and Cognition, 28,* 748–769. https://doi.org/10.1037//0278-7393.28.4.748

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. https://doi.org/10.18637/jss.v067.i01

Bowles, N. L., & Poon, L. W. (1981). The effect of age on speed of lexical access. *Experimental Aging Research, 7,* 417–425. https://doi.org/10.1080/03610738108259822

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology, 40,* 71–77. https://doi.org/10.1093/geronj/40.1.71

Chien, Y. C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1,* 225–295.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, Netherlands: Foris.

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5,* 630. https://doi.org/10.3389/fpsyg.2014.00630

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language, 65,* 128–144.

Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words him and her: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language, 36,* 276–292. https://doi.org/10.1006/jmla.1996.2499

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 59,* P203–P209. https://doi.org/10.1093/geronb/59.5.P203

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes, 28,* 188–219. https://doi.org/10.1080/01690965.2010.548391

Cunnings, I., Patterson, C., & Felser, C. (2015). Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology, 6,* 840. https://doi.org/10.3389/fpsyg.2015.00840

Davis, S. W., Zhuang, J., Wright, P., & Tyler, L. K. (2014). Age-related sensitivity to task-related modulation of language-processing networks. *Neuropsychologia, 63,* 107–115. https://doi.org/10.1016/j.neuropsychologia.2014.08.017

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4,* 500–503. https://doi.org/10.1037/0882-7974.4.4.500

Feyereisen, P. (1997). A meta-analytic procedure shows an age-related decline in picture naming: Comments on Goulet, Ska, and Kahn (1994). *Journal of Speech, Language, and Hearing Research, 40,* 1328–1333. https://doi.org/10.1044/jslhr.4006.1328

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Grodzinsky, Y., Wexler, K., Chien, Y. C., Marakovitz, S., & Solomon, J. (1993). The breakdown of binding relations. *Brain and Language, 45,* 396–422. https://doi.org/10.1006/brln.1993.1052

Grossman, M., Cooke, A., DeVita, C., Alsop, D., Detre, J., Chen, W., & Gee, J. (2002). Age-related changes in working memory during sentence comprehension: an fMRI study. *Neuroimage, 15,* 302–317. https://doi.org/10.1006/nimg.2001.0971

Guasti, M. (2002). *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26,* 433–443. https://doi.org/10.1177/0956797614567339

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press. https://doi.org/10.1016/S0079-7421(08)60041-9

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.

Hoogendam, Y. Y., Hofman, A., van der Geest, J. N., van der Lugt, A., & Ikram, M. A. (2014). Patterns of cognitive function in aging: The Rotterdam Study. *European Journal of Epidemiology, 29,* 133–140. https://doi.org/10.1007/s10654-014-9885-4

Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25,* 1–44.

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics, 7,* 277–287. https://doi.org/10.1017/S0142716400007578

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging, 12,* 362–371. https://doi.org/10.1037//0882-7974.12.2.362

Kennison, S. M. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language, 49,* 335–352. https://doi.org/10.1016/S0749-596X(03)00071-8

Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics, 36,* 1317–1374.

Koster, J. (1993). *Toward a new theory of anaphoric binding* (Doctoral dissertation). The Netherlands: University of Groningen.

Leonard, C. L., Waters, G. S., & Caplan, D. (1997). The influence of contextual information on the resolution of ambiguous pronouns by younger and older adults. *Applied Psycholinguistics, 18,* 293–317. https://doi.org/10.1017/S0142716400010493

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29,* 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10,* 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology, 40,* 737–747. https://doi.org/10.1093/geronj/40.6.737

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

Manard, M., Carabin, D., Jaspar, M., & Collette, F. (2014). Age-related decline in cognitive control: The role of fluid intelligence and processing speed. *BMC Neuroscience, 15,* 7. https://doi.org/10.1186/1471-2202-15-7

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29,* 111–123. https://doi.org/10.1023/A:1005184709695

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48,* 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech, 48,* 123–156. https://doi.org/10.1177/00238309050480020101

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18,* 5–19. https://doi.org/10.1007/BF01069043

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's

area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 263–281. https://doi.org/10.3758/CABN.5.3.263

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology, 5,* 147. https://doi.org/10.3389/fpsyg.2014.00147

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://R-project.org

Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: Evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition, 24,* 672–702. https://doi.org/10.1080/13825585.2016.1251550

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience, 29,* 912–927.

Ruigendijk, E., Vasić, N., & Avrutin, S. (2006). Reference assignment: Using language breakdown to choose between theoretical approaches. *Brain and Language, 96,* 302–317. https://doi.org/10.1016/j.bandl.2005.06.005

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: The network dynamics of cognitive decline and preservation. *Science, 346*(6209), 583–587. https://doi.org/10.1126/science.1254404

Shake, M. C., & Stine-Morrow, E. A. (2011). Age differences in resolving anaphoric expressions during reading. *Aging, Neuropsychology, and Cognition, 18,* 678–707. https://doi.org/10.1080/13825585.2011.607228

Sigurjónsdóttir, S., & Hyams, N. (1992). Reflexivization and logophoricity: Evidence from the acquisition of Icelandic. *Language Acquisition, 2,* 359–413.

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: The modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex, 20,* 352–364. https://doi.org/10.1093/cercor/bhp105

Ueno, M., & Kehler, A. (2010). The interpretation of null and overt pronouns in Japanese: Grammatical and pragmatic factors. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2057–2062). Austin, TX: Cognitive Science Society.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122,* 231–249. https://doi.org/10.1037/0033-2909.122.3.231

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, 49,* 51–74. https://doi.org/10.1080/713755607

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16,* 128–144. https://doi.org/10.1037/0882-7974.16.1.128

Appendix A (p. 1 of 2)

Experimental sentences (and gloss) for Experiment 2. Experimental items in Experiment 1 were identical except that the adjective preceding NP2 and the PP following NP2 were omitted.

1. *Dennis/Sarah beschließt, dass der/die talentierte Journalist/Journalistin aus Athen ihn interviewen soll.*
   Dennis/Sarah decides that the talented journalist from Athens should interview him.
2. *Georg/Annika bittet, dass der/die reiche Onkel/Tante aus Luxemburg ihn morgen abholt.*
   Georg/Annika asks that the rich uncle/aunt from Luxembourg fetches him tomorrow.
3. *Benjamin/Elisabeth sieht, dass der/die begabte Tänzer/Tänzerin aus Argentinien ihn dauernd anlächelt.*
   Benjamin/Elisabeth sees that the talented dancer from Argentina smiles at him constantly.
4. *Philipp/Charlotte verlangt, dass der/der freche Kassierer/Kassiererin aus Dortmund ihn freundlicher behandelt.*
   Philipp/Charlotte demands that the cheeky cashier from Dortmund treats him better.
5. *Jonathan/Jennifer ruft, dass der/die blonde Lehrer/Lehrerin aus Bremen ihn beachten soll.*
   Jonathan/Jennifer shouts that the blond teacher from Bremen should notice him.
6. *Richard/Sophia ahnt, dass der/die ungeduldige Polizist/Polizistin aus Bochum ihm nicht glaubt.*
   Richard/Sophia suspects that the impatient policeman/policewoman from Bochum doesn't believe him.
7. *Frank/Linda mag, dass der/die unerfahrene Sekretär/Sekretärin aus Münster ihn sehr bewundert.*
   Frank/Linda likes that the inexperienced secretary from Münster admires him a lot.
8. *Jacob/Emma erkennt, dass der/die fleißige Student/Studentin aus Korea ihm sehr vertraut.*
   Jacob/Emma realizes that the hard-working student from Korea trusts him a lot.
9. *Christian/Victoria verspricht, dass der/die geliebte Enkel/Enkelin aus Nürnberg ihn besuchen darf.*
   Christian/Victoria promises that the beloved grandson/granddaughter from Nuremberg may visit him.
10. *Max/Julia verhindert, dass der/die bekannte Richter/Richterin aus Karlsruhe ihn verurteilen kann.*
    Max/Julia prevents that the well-known judge from Karlsruhe can convict him.
11. *Martin/Barbara erwähnt, dass der/die hübsche Prinz/Prinzessin aus Spanien ihm zugewinkt hat.*
    Martin/Barbara mentions that the handsome prince from Spain waved at him.
12. *Patrick/Stephanie spürt, dass der/die alte Arzt/Ärztin aus Japan ihn ernst nimmt.*
    Patrick/Stephanie senses that the old doctor from Japan takes him seriously.
13. *Simon/Laura betont, dass der/die junge Beamte/Beamtin aus Kreuzberg ihm geholfen hat.*
    Simon/Laura emphasizes that the young clerk from Kreuzberg has helped him.
14. *Erik/Jasmin verkündet, dass der/die liebe Opa/Oma aus München ihm geschrieben hat.*
    Erik/Jasmin announces that the nice grandpa/grandma from Munich has written him.
15. *Martin/Barbara fürchtet, dass der/die schnelle Dieb/Diebin aus Frankfurt ihm entkommen ist.*
    Martin/Barbara fears that the fast thief from Frankfurt escaped him.
16. *Paul/Clara lobt, dass der/die stille Schüler/Schülerin aus Russland ihn beeindruckt hat.*
    Paul/Clara praises that the quiet student from Russia has impressed him.
17. *Max/Julia sagt, dass der/die große Künstler/Künstlerin aus Brasilien ihn malen soll.*
    Max/Julia says that the great artist from Brazil should paint him.
18. *Michael/Carmen bemerkt, dass der/die aufgeregte Koch/Köchin aus China ihn vergessen hat.*
    Michael/Carmen notices that the nervous cook from China has forgotten him.
19. *David/Anna hört, dass der/die neue Kollege/Kollegin aus Schweden ihn nicht mag.*
    David/Anna hears that the new colleague from Sweden doesn't like him.
20. *Thomas/Alexandra genießt, dass der/die erfolgreiche Musiker/Musikerin aus Finnland ihm etwas vorspielt.*
    Thomas/Alexandra enjoys that the successful musician from Finland plays him something.
21. *Lukas/Natalie vergisst, dass der/die hektische Reporter/Reporterin aus Hamburg ihn falsch zitierte.*
    Lukas/Natalie forgets that the hectic reporter from Hamburg misquoted him.
22. *Oscar/Josephine glaubt, dass der/die kleine Neffe/Nichte aus Frankreich ihn besuchen wird.*
    Oscar/Josephine thinks that the little nephew/niece from France will visit him.
23. *Christian/Victoria erzählt, dass der/die nette Mönch/Nonne aus Freiburg ihm viel erklärte.*
    Christian/Victoria says that the nice monk/nun from Freiburg explained a lot to him.
24. *Benjamin/Elisabeth bestätigt, dass der/die unehrliche Politiker/Politikerin aus Mainz ihn unglücklich macht.*
    Benjamin/Elisabeth confirms that the dishonest politician from Mainz makes him unhappy.

Appendix A (p. 2 of 2)

Experimental sentences (and gloss) for Experiment 2. Experimental items in Experiment 1 were identical except that the adjective preceding NP2 and the PP following NP2 were omitted.

25. *Daniel/Jessica will, dass der/die gestresste Apotheker/Apothekerin aus Wien ihn sofort bedient.*
Daniel/Jessica wants that the stressed pharmacist from Vienna serves him first.

26. *Robert/Diana erfährt, dass der/die fröhliche Bäcker/Bäckerin aus Schöneberg ihn überraschen wollte.*
Robert/Diana learns that the happy baker from Schöneberg wanted to surprise him.

27. *Erik/Jasmin verrät, dass der/die reiche Graf/Gräfin aus Monaco ihn betrogen hat.*
Erik/Jasmin reveals that the rich count/countess from Monaco has betrayed him.

28. *Jonathan/Jennifer äußert, dass der/die nervöse Übersetzer/Übersetzerin aus Bielefeld ihm komisch vorkommt.*
Jonathan/Jennifer remarks that the nervous translator from Bielefeld strikes him as odd.

29. *Georg/Annika liest, dass der/die freundliche Gärtner/Gärtnerin aus Italien ihn einstellen möchte.*
Georg/Annika reads that the friendly gardener from Italy wants to employ him.

30. *Frank/Linda meldet, dass der/die verletzte Sportler/Sportlerin aus Bonn ihn bestechen wollte.*
Frank/Linda reports that the injured athlete from Bonn wanted to bribe him.

31. *Oliver/Hanna fordert, dass der/die unsichere Praktikant/Praktikantin aus Potsdam ihm mehr hilft.*
Oliver/Hanna demands that the insecure intern from Potsdam helps him more.

32. *Sebastian/Tina bezweifelt, dass der/die hilfsbereite Anwalt/Anwältin aus Düsseldorf ihn beraten wird.*
Sebastian/Tina doubts that the helpful lawyer from Düsseldorf will counsel him.

33. *Michael/Carmen akzeptiert, dass der/die fleißige Schneider/Schneiderin aus Mannheim ihn warten lässt.*
Michael/Carmen accepts that the hard-working tailor from Mannheim makes him wait.

34. *Robert/Diana findet, dass der/die erfahrene Forscher/Forscherin aus Dänemark ihm helfen sollte.*
Robert/Diana thinks that the experienced researcher from Denmark should help him.

35. *Daniel/Jessica erklärt, dass der/die verwirrte Dolmetscher/Dolmetscherin aus Brüssel ihn nicht versteht.*
Daniel/Jessica explains that the confused interpreter from Brussels doesn't understand him.

36. *Patrick/Stephanie hofft, dass der/die kluge Therapeut/Therapeutin aus Österreich ihm helfen kann.*
Patrick/Stephanie hopes that the smart therapist from Austria can help him.

37. *Dennis/Sarah bestimmt, dass der/die schüchterne Assistent/Assistentin aus Amsterdam ihn morgen vertritt.*
Dennis/Sarah decides that the shy assistant from Amsterdam fills in for him tomorrow.

38. *David/Anna verdient, dass der/die launische Berater/Beraterin aus Wuppertal ihn mehr unterstützt.*
David/Anna deserves that the moody consultant from Wuppertal supports him more.

39. *Oscar/Josephine merkt, dass der/die beschäftigte Kellner/Kellnerin aus Mexiko ihn nicht beachtet.*
Oscar/Josephine notices that the busy waiter/waitress from Mexico doesn't pay attention to him.

40. *Philipp/Charlotte träumt, dass der/die berühmte Schauspieler/Schauspielerin aus Amerika ihn endlich küsst.*
Philipp/Charlotte dreams that the famous actor/actress from America finally kisses him.

41. *Simon/Laura begreift, dass der/die lustige Redakteur/Redakteurin aus Mainz ihn verunsichern will.*
Simon/Laura realizes that the funny editor from Mainz wants to unnerve him.

42. *Sebastian/Tina drängt, dass der/die ängstliche Bauer/Bäuerin aus Brandenburg ihn anrufen soll.*
Sebastian/Tina urges that the fearful farmer from Brandenburg should call him.

43. *Oliver/Hanna denkt, dass der/die arrogante Professor/Professorin aus Belgien ihm nicht zuhört.*
Oliver/Hanna thinks that the arrogant professor from Belgium doesn't listen to him.

44. *Lukas/Natalie beweist, dass der/die unfreundliche Verkäufer/Verkäuferin aus Berlin ihn belogen hat.*
Lukas/Natalie proves that the unfriendly salesman/saleswoman from Berlin has lied to him.

45. *Jacob/Emma erwartet, dass der/die kreative Friseur/Friseurin aus Paris ihn gut berät.*
Jacob/Emma expects that the creative hairdresser from Paris advises him well.

46. *Richard/Sophia weiß, dass der/die glückliche Autor/Autorin aus Brasilien ihm danken möchte.*
Richard/Sophia knows that the happy author from Brazil wants to thank him.

47. *Thomas/Alexandra behauptet, dass der/die unhöfliche Moderator/Moderatorin aus Köln ihn ständig unterbricht.*
Thomas/Alexandra claims that the impolite host from Cologne interrupts him constantly.

48. *Paul/Clara fühlt, dass der/die strenge Erzieher/Erzieherin aus Dresden ihn oft ignoriert.*
Paul/Clara feels that the strict teacher from Dresden often ignores him.

Appendix B (p. 1 of 3)

Table B1. Results of Experiment 1, generalized linear mixed-effects model for acceptability rates. (Note that because none of the interactions with working memory [WM] score reached significance, these values were omitted from the results table for the sake of brevity.)

Fixed effects	β	*SE*	*z* value	*p* value
Intercept	2.266	0.175	12.96	< .001
NP1-gender	2.210	0.140	15.74	< .001
NP2-gender	0.049	0.151	0.33	.744
NP2-frequency	0.010	0.043	0.24	.813
NP2-length	–0.012	0.037	–0.34	.736
Age group	–0.173	0.332	–0.52	.603
Trial number	0.010	0.004	2.47	.014
WM score	0.030	0.060	0.51	.609
NP1-gender:age group	0.615	0.274	2.25	.025
NP2-gender:age group	0.310	0.233	1.33	.184
NP1-gender:NP2-gender	0.012	0.037	0.34	.735
NP1-gender:NP2-gender:age group	0.325	0.545	0.60	.551

Random effects		Variance	*SD*	Correlation
Items	Intercept	1.296	1.139	
	Age group	0.105	0.323	–.60
Participants	Intercept	0.001	0.024	

NP1 = noun phrase 1; NP2 = noun phrase 2.

Table B2. Results of Experiment 1, linear mixed-effects model for judgment response times. (Note that because none of the interactions with working memory [WM] score reached significance, these values were omitted from the results table for the sake of brevity.)

Fixed effects	β	*SE*	*t* value
Intercept	–1.582	0.069	–23.06
NP2-gender	–0.017	0.025	–0.65
NP2 gender	0.004	0.030	0.13
NP2-frequency	0.006	0.008	0.81
NP2-length	0.001	0.006	0.11
Age group	–0.552	0.137	–4.03
Trial number	–0.007	0.001	–8.15
WM score	–0.008	0.026	0.30
NP1-gender:age	–0.100	0.051	–1.97
NP2-gender:age	0.002	0.050	0.03
NP1-gender:NP2-gender	–0.008	0.050	–0.16
NP1-gender:NP2-gender:age group	0.036	0.010	0.36

Random effects		Variance	*SD*	Correlation
Items	Intercept	0.272	0.522	
	Age group	0.011	0.103	–.97
Participants	Intercept	0.0001	0.009	
Residual		0.340	0.583	

NP1 = noun phrase 1; NP2 = noun phrase 2.

Table B3. Overview of significant effects on self-paced reading times, Experiment 1.

Position and word	Effect	β	*SE*	*t* value
1 *Erik/Jasmin*	Age	–0.818	0.112	–7.28
	Trial number	–0.007	0.001	–11.92
2 *verkündet*	Age	–0.824	0.119	–6.92
	Trial number	–0.007	0.001	–10.30
3 *dass*	Age	–0.775	0.103	–7.50
	Trial number	–0.007	0.001	–12.57
4 *der/der*	Age	–0.889	0.108	–8.24
	Trial number	–0.006	0.001	–11.97
5 *Opa/Oma*	Age	–0.901	0.114	–7.87
	Trial number	–0.006	0.001	–10.01
	NP2-frequency	–0.010	0.005	–2.16
	NP2-length	0.031	0.004	7.27
6 *ihm*	Age	–0.992	0.098	–10.08
	Trial number	–0.004	0.001	–5.90
	NP1-gender	–0.103	0.017	–5.90
	NP2-frequency	–0.013	0.005	–2.40
	NP2-length	0.010	0.004	2.20
	NP1-gender:age group	0.088	0.035	2.53
7 *geschrieben*	Age	–0.654	0.084	–7.82
	Trial number	–0.007	0.001	–10.87
	NP1-gender	0.057	0.018	3.17
	NP2-length	–0.014	0.005	–2.61
	NP1-gender:age group	–0.185	0.036	–5.17

NP1 = noun phrase 1; NP2 = noun phrase 2.

Table B4. Results of Experiment 2, generalized linear mixed-effects model for acceptability rates.

Fixed effects	β	*SE*	*z* value	*p* value
Intercept	1.34	0.20	6.76	< .001
NP1-gender	4.17	0.16	25.72	< .001
NP2-gender	0.23	0.18	1.31	.191
NP2-frequency	0.05	0.04	1.22	.221
NP2-length	–0.001	0.04	–0.02	.983
Age group	–0.22	0.38	–0.57	.572
Trial number	0.01	0.003	2.98	.003
NP1-gender:age group	0.89	0.31	2.81	.005
NP2-gender:age group	–0.05	0.27	–0.20	.844
NP1-gender:NP2-gender	0.28	0.27	1.05	.292
NP1-gender:NP2-gender:age group	0.46	0.54	0.84	.399
Random effects		**Variance**	***SD***	**Correlation**
Items	Intercept	0.169	0.411	
	Age group	0.222	0.471	–.47
Participants	Intercept	1.806	1.344	

NP1 = noun phrase 1; NP2 = noun phrase 2.

Table B5. Results of Experiment 2, linear mixed-effects model for response time.

Fixed effects		β	*SE*	*t* value
Intercept		–3.23	0.34	–9.59
NP1-gender		1.18	0.50	2.36
NP2-gender		1.90	0.65	2.93
NP2-frequency		–0.08	0.16	–0.53
NP2-length		0.36	0.14	2.55
Age group		–0.42	0.10	–4.02
Trial number		–0.01	0.01	–1.39
NP1-gender:age group		–1.63	1.00	–1.63
NP2-gender:age group		–3.25	1.29	–2.53
NP2-length:age group		–0.72	0.28	–2.59
NP1-gender:NP2-gender		–2.09	0.99	–2.11
NP1-gender:NP2-gender:age group		4.70	1.99	2.37
Random effects		**Variance**	***SD***	**Correlation**
Items	Intercept	1.54	1.24	
	Age group	5.65	2.38	–.62
Participants	Intercept	1.41	1.19	
Residual		128.95	11.36	

NP1 = noun phrase 1; NP2 = noun phrase 2.