American Psychological Association

Psychology and Aging

0882-7974/19/$12.00

2019, Vol. 34, No. 6, 791–804
http://dx.doi.org/10.1037/pag0000372

Age-Related Differences in Referential Production: A Multiple-Measures Study

Raheleh Saryazdi
University of Toronto

Julie Bannon
McMaster University

Craig G. Chambers
University of Toronto

Contemporary research on aging has provided mixed evidence for whether older adults are less effective than younger adults at designing and delivering spoken utterances. However, most of these studies have focused on only specific aspects of this process. In addition, they tend to vary significantly in terms of the degree of complexity in their chosen stimuli or task. The present study compares younger and older adults' performance using a referential production paradigm involving simple everyday objects. We varied referential context such that a target object was either unique in its category (e.g., one shirt), or was accompanied by a same-category object (e.g., two shirts). We evaluated whether speakers' descriptions provided listeners with sufficient information for identification, and whether speakers spontaneously adapt their speech for different addressee types (younger adult, older adult, automated dialogue system). A variety of measures were included to provide a comprehensive perspective on adults' performance. Interestingly, the results revealed few or no age differences in measures related to production performance (speech onset latency, speech rate, and fluency). In contrast, consistent differences were observed for measures related to descriptive content, both in terms of informativity and variability in lexical selection: Older adults not only provided more information than necessary for referential success (e.g., superfluous modifiers), but also exhibited greater variability in their selection of modifiers. The results show that, although certain aspects of the production process are well-preserved across the adult lifespan, meaningful age-related differences can still be found in simple referential tasks with everyday objects.

Keywords: referential production, aging, informativity, audience design, speaking performance

Supplemental materials: http://dx.doi.org/10.1037/pag0000372.supp

Communication is a key element in healthy aging in view of its importance for maintaining social relationships, participating in meaningful activities, and performing the informational exchanges involved in everyday events. Intriguingly, language abilities are often described as comparatively resilient to effects of age even though these abilities are supported by an intricate interplay of perceptual, cognitive, and motor systems (Burke & Shafto, 2008; Harada, Natelson Love, & Triebel, 2013; Hartshorne & Germine, 2015; Shafto & Tyler, 2014). However, meaningful differences can be identified when the focus is narrowed to specific linguistic processes and when sufficiently sensitive measures are used. Here, we explore potential age differences in referential production, with a particular focus on speakers' ability to generate descriptions of ordinary objects. In addition to measuring various aspects of speaking performance, we compare the ability of older and younger adults to tailor the content of their descriptions according to the informational requirements of the context, as well as the type of addressee (older adult, younger adult, automated dialogue system).

The act of speaking seems deceptively simple despite the range of psycholinguistic processes involved in the conceptualization, planning, and execution of spoken utterances (see Levelt, 1989). One important issue concerns the stability of these processes over the adult lifespan. In light of the various age-related declines that occur in sensory-perceptual processing, cognition, and motor behavior, exactly how effective are older adults at generating speech in comparison to their younger counterparts? This question can be explored in relation to both the "performance" aspects of older adults' speech (considerations of timing and fluency) and the

This article was published Online First June 17, 2019.

Raheleh Saryazdi, Department of Psychology, University of Toronto; Julie Bannon, Department of Psychology, Neuroscience, and Behaviour, McMaster University; Craig G. Chambers, Department of Psychology, University of Toronto.

Funding for this research was provided by a fellowship to Raheleh Saryazdi and a grant to Craig G. Chambers from the Natural Sciences and Engineering Research Council (Canada). The authors have no conflicts of interest to declare. We thank Lena Donald, Chris Klammer, Yvette Hou, and Joanne Nuque for their assistance with this research. Selected results from the current study were presented at the 2017 Architectures and Mechanisms for Language Processing Conference, the 2018 CUNY Sentence Processing Conference, and the 2018 Lake Ontario Visionary Establishment Conference. The authors will provide raw data and visual materials upon request.

Correspondence concerning this article should be addressed to Raheleh Saryazdi, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada. E-mail: raheleh.saryazdi@mail.utoronto.ca

content of speech (the relevance and sufficiency of the information being communicated).

Performance Measures of Language Production

To date, empirical studies have often reported less efficient patterns of performance in older adults' delivery of spoken utterances (see Mortensen, Meyer, & Humphreys, 2006 for a review). For example, older adults tend to be slower at initiating speech (e.g., Mortensen, Meyer, & Humphreys, 2008; Spieler & Griffin, 2006), a measure that is taken to reflect the conceptualization processes involved in planning an utterance (Griffin, 2003; Griffin & Bock, 2000). Slower speech onset latencies are also sometimes attributed to a specific difficulty in lexical retrieval, given evidence that older adults experience more tip-of-the-tongue states than younger adults (e.g., Burke & Shafto, 2004; Griffin & Spieler, 2006; Shafto, Burke, Stamatakis, Tam, & Tyler, 2007). However, although difficulty with word retrieval can be attributed to the general age-related decline in cognitive processes, it could also arise as a result of greater competition during word selection due to older adults' larger vocabulary (Griffin & Spieler, 2006; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014, but see Shafto, James, Abrams, & Tyler, 2017). In fact, when comparing tip-of-the-tongue phenomena across age groups, age differences may disappear when vocabulary size is controlled (e.g., Dahlgren, 1998).

Difficulty with word selection may also explain older adults' reported tendency to continue to speak at a slower rate during utterance delivery (e.g., Duchin & Mysak, 1987; Horton, Spieler, & Shriberg, 2010; Ramig, 1983; Smith, Wasowicz, & Preston, 1987; Verhoeven, De Pauw, & Kloots, 2004). However, this behavior has also been suggested to reflect a compensatory strategy whereby older adults engage in more careful utterance planning to reduce their concurrent processing load and maintain the fluency of speech (Belke & Meyer, 2007; Mortensen et al., 2008). On this account, a trade-off situation ensues: More time spent preparing words prior to articulation leads to a lower incidence of disfluency, particularly in older populations.

Disfluency itself is one of the most studied aspects of speech behavior in the aging and language production literature. Disfluencies are quite common in natural discourse, with reports suggesting six out of 100 words to be disfluent (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). Because the nature of language production is considered to be incremental (Levelt, 1989; Pechmann, 1989), the presence of disfluency during production is often considered to be a sign of difficulty with speech planning (Bortfeld et al., 2001; Clark & Wasow, 1998; Fox Tree, 1995; Fraundorf & Watson, 2013; Mortensen et al., 2008; Oviatt, 1995; Watanabe, Hirose, Den, & Minematsu, 2008). Increased disfluency could therefore arise from greater difficulty in finding words (Hartsuiker & Notebaert, 2010; James, Chambers, & Placzek, 2018; Kavé & Goral, 2017). Although some studies have indeed reported more disfluency in older adults (e.g., Bortfeld et al., 2001; Horton et al., 2010; James et al., 2018; Mortensen et al., 2008), others have failed to show any age-related differences (e.g., Duchin & Mysak, 1987; Kemper, Ferrell, Harden, Finter-Urczyk, & Billington, 1998). One goal of the current study is to provide additional evidence for the timing and fluency features of older adults' speech using a simple yet naturalistic reference-production task.

Measures of Descriptive Content

In addition to the measures described above, the assessment of language production abilities must take into account whether speakers generate utterances that contain appropriate communicative content. Psycholinguistic studies have often used the production of referential descriptions in particular as a tool to explore speakers' ability to calculate the informational needs of a listener in relation to the situational context. According to Grice's (1975) Maxim of Quantity, adequate referential descriptions should contain just the right amount of information to successfully identify a referent—neither too little nor too much. The ability to tailor a referential description according to the situation is often studied using paradigms where the number of candidate referents is varied. For example, a visual scene might contain, among other things, two distinct shirts, one of which is the intended referent. In order for a listener to successfully identify the target, the speaker must provide specific information (e.g., *the dress shirt, the shirt on the right*) relative to what could be used otherwise (e.g., *the shirt*). Although younger adults do not always provide the amount of information that is consistent with a strict interpretation of Gricean maxims (e.g., Engelhardt, Bailey, & Ferreira, 2006; Rubio-Fernandez, 2016; Tarenskeen, Broersma, & Geurts, 2015), they seem more likely to do so in comparison to older adults (Healey & Grossman, 2016; Long, Horton, Rohde, & Sorace, 2018). For instance, using a task in which display objects contrasted in terms of multiple perceptual properties, Healey and Grossman (2016) showed that older adults were more likely than younger adults to provide descriptions that contained either too much or too little information.

An additional question concerns the specific types of properties that speakers encode in their referential descriptions. For example, speakers often show a strong tendency to use color modifiers when possible, which may arise from the intrinsic and salient nature of color features, in turn boosting the accessibility of this information for linguistic reference (Belke & Meyer, 2002; Fukumura, 2018; Heller & Chambers, 2014). Indeed, overspecified descriptions often contain color terms, again reflecting the salience of this property (Tarenskeen et al., 2015). Further, in visual contexts, color overspecification could facilitate referential comprehension because of the way in which color rapidly narrows the search space for a listener (Rubio-Fernandez, 2016). A related hypothesis is that overspecification (using any type of modifier) could reflect a strategy for managing the demands of online utterance formulation. More specifically, speakers might add a redundant modifier (e.g., *the white shirt*) to "buy time" while accessing a noun for the referential target (Pechmann, 1989). Such an explanation would not apply, however, to the redundant use of modifiers that occur postnominally (e.g., *the shirt on the right side*), where the noun had been successfully selected before the modifier occurred. Psycholinguistic accounts of referential production are therefore informed by considering the type of modifiers selected by speakers, as well as whether they occur prenominally or postnominally.

A phenomenon that shares features with the process of formulating an appropriately informative description is the ability to adjust one's speech according to characteristics of the specific addressee (often referred to as audience design, see Clark & Murphy, 1982; Coupland & Giles, 1988; Horton & Gerrig, 2002). The most intuitive example of this is the way that speakers adjust their speech for young children (infant-/child-directed speech; Fernald & Simon, 1984; Newman-Norlund et al., 2009; Uther, Knoll, & Burnham, 2007). However, accommodations also occur when speakers are addressing older adults. This can involve increases in the number of words and utterances, simplifications in syntactic structure, and performance differences of the type mentioned earlier, such as slower speaking rate (Giles & Gasiorek, 2011; Hummert, Shaner, Garstka, & Henry, 1998; Kemper, 1994; Kemper et al., 1998; Kemper & Harden, 1999; Kemper, Othick, Warren, Gubarchuk, & Gerhing, 1996). In the primary literature, this is often referred to as "elderspeak" (Kemper, 1994). Further, an emerging research theme inspired by advances in technology is how speakers make adaptations when directing speech to automated systems equipped with spoken language interfaces. Studies in this area have suggested that speakers may hyperarticulate, speak slower, and produce few disfluencies when directing speech to a computer than to another human (e.g., Bell, Gustafson, & Heldner, 2003; Oviatt, 1995; Oviatt, Bernard, & Levow, 1998; Oviatt, MacEachern, & Levow, 1998; Stent, Huffman, & Brennan, 2008). In terms of descriptive content, speakers tend to also vary their referential expressions to match the processing capability of a computer addressee (e.g., Bell et al., 2003; Branigan, Pickering, Pearson, McLean, & Brown, 2011; Schmader & Horton, 2019). A number of studies, however, have failed to find evidence of adaptation in speech directed to computers (e.g., Amalberti, Carbonell, & Falzon, 1993; Bergmann, Branigan, & Kopp, 2015; van Lierop, Goudbeek, & Krahmer, 2012). The current study applies a multiple-measures approach to explore adaptations in speech to both humans and computers, and the extent to which any observed patterns might vary across the adult lifespan.

To what extent do older adults adapt their speech for their addressees? To date, research on this topic (involving human addressees) suggests that, in comparison to younger adults, older adults are less likely to make accommodations. For example, older speakers are less likely to adjust speech according to whether it is directed to a younger or an older addressee (Kemper & Harden, 1999; Kemper et al., 1996; Kemper, Vandepute, Rice, Cheung, & Gubarchuk, 1995; Schubotz, Özyürek, & Holler, 2019), or adjust the amount of information needed to communicate with a knowledgeable versus naïve conversational partner (Horton & Spieler, 2007, but see Yoon & Stine-Morrow, 2019). Older adults, however, do accommodate their speech when communicating with a listener who has a cognitive impairment (in contrast to a healthy listener; Gould & Shaleen, 1999; Kemper et al., 1998). Interestingly, the degree to which older adults adjust their speech for an addressee appears to be modulated by task demands, with more cognitively demanding tasks leading to less accommodation (Gould & Shaleen, 1999). This general idea has in fact been echoed in research on language comprehension. For example, Campbell et al. (2016) argued that task difficulty can determine whether age differences are found in listening tasks. To minimize task demands, the present study employed a game-like methodology with images of everyday objects.

Present Study

Studies exploring referential production across the lifespan have typically focused on specific aspects of the production system in isolation (e.g., speech onset latencies, disfluency, or informational content). This limits the ability to develop a more comprehensive understanding of older adults' relative success at producing utterances that are well timed, fluent, and informationally sufficient. Another limitation is that some studies have used paradigms where speakers produced only single words (object naming), reducing the time frame in which it is possible to explore the range of effects. In the current study, we employed a task in which full sentences are elicited, and where numerous measures related to both language performance and informational content are collected simultaneously. We used a referential communication paradigm in which participants were asked to provide instructions for an addressee (a methodology that has been fruitful for exploring various aspects of language production). To date, studies of aging that have used this paradigm have often employed complex visual stimuli such as unfamiliar figures (e.g., tangrams; Bortfeld et al., 2001; Hupet, Chantraine, & Nef, 1993; Lysander & Horton, 2012) or stimulus sets designed to have high visual similarity across tokens (e.g., different kinds of birds; Horton & Spieler, 2007). This is an important consideration given independent evidence that task complexity can have a significant influence on measures such as speech onset latencies (e.g., Davies & Kreysa, 2017; Elsner, Clarke, & Rohde, 2018; Gatt, Krahmer, van Deemter, & van Gompel, 2017), speech rate (e.g., Duchin & Mysak, 1987), disfluency (e.g., Brown-Schmidt & Tanenhaus, 2006; Schachter, Christenfeld, Ravina, & Bilous, 1991), and lexical content (e.g., Davies & Katsos, 2013; Elsner et al., 2018; Healey & Grossman, 2016).

The present task, in contrast, involved an array of four objects that people would routinely see and refer to in daily life. Within this array, the object serving as the target was either unique in its category (e.g., one dog), or was accompanied by another exemplar in the same category (a second dog), reflecting a well-established manipulation for evaluating how speakers vary the informativity of their descriptions according to the context. Beyond the issue of whether speakers are successful at adjusting informativity (e.g., saying the *spotted dog* instead of simply the *dog* when two dogs are present), we also measured the type of modifier(s) and whether the modifier(s) occurred prenominally or postnominally. These measures can provide additional insight into the question of whether referential production mechanisms differ for younger and older adults.

Another feature of the current study involved the ability to capture different lexical strategies that a speaker might use to ensure successful reference. An assumption reflected in the methodology of many past studies is that listeners need to add modifying information to ensure a description is successful in the presence of contrastive objects. However, in natural language, subordinate terms can often be used to achieve referential success (e.g., the term *Dalmatian* might be sufficient to differentiate an intended dog from another in the context). Despite the availability of this option in many natural contexts, there has been little discussion of how listeners might employ subordinates in referential production tasks (although see Brennan & Clark, 1996; Yoon & Brown-Schmidt, 2013). To address this issue, the present study included target stimuli that could be differentiated either using a modified basic-level noun or a subordinate term, allowing further exploration of how speakers manage the content of their descriptions across the adult lifespan. Here, we consider basic-level nouns to be object labels that correspond to basic taxonomic categories (e.g., *dog* as opposed to *animal*) as discussed by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976).

Finally, we also examined speakers' ability to tailor utterances according to the type of individual (audience design). This involved a manipulation in which participants were asked to provide descriptions for three different imagined types of addressees: younger adult, older adult, or an automated dialogue system.

Method

Participants

The younger adult group consisted of 24 students (age range = 18 to 23, M = 19.38, SD = 1.38, 20 female) recruited from a psychology program participant pool at the University of Toronto Mississauga. These participants received partial course credit or $10 per hour as compensation. The older adult group consisted of 24 volunteers (age range = 69 to 83, M = 75.33, SD = 3.63, 14 female) recruited from the University of Toronto Mississauga senior database pool and received $15 per hour as compensation. All participants reported normal hearing, and normal or corrected-to-normal vision. In addition, older adults were screened for hearing, visual acuity, and color blindness. All participants reported native or native-like competence in English. More specifically, the majority of participants reported learning English prior to the age of 5 (92% of younger and 92% of older adults) with the remaining learning English by the age of 10. In addition, younger and older adults reported "very fluent" or "fluent" English language proficiency on the language questionnaire. Both age groups were also quite comparable in terms of their educational background: On average, older adults completed high school plus two years of higher education, whereas younger adults completed high school plus one year of higher education. As part of the requirement for joining the senior database pool, all older adults completed an abbreviated Mill Hill vocabulary test (Raven, 1965) whereby they had to select the best synonym for a given word. The average score was 15.88/20 (SD = 1.86). This was near identical to scores reported in a separate study conducted at the same site comparing younger and older adults by Ben-David, Erel, Goy, and Schneider (2015). In that study, the average score for older adults was 15. Although we did not collect Mill Hill scores for our younger participants in the current study, we would expect them to be comparable to Ben-David et al. (2015) as they were drawn from the same undergraduate population (M = 13.2). The study protocol was approved by the Social Sciences, Humanities, and Education Research Ethics Board at the University of Toronto (Number 30335).

Materials

The present study employed a variation of a referential communication task in which participants had to provide a series of task instructions for three different types of addressees, namely a younger adult, an older adult, and an automated dialogue system (hereafter referred to as "computer"). There were 12 trials for each of the three addressee types, six of which were critical trials, and six of which were filler trials, yielding a total of 18 critical trials per participant. The trials corresponding to a given addressee type were presented together in a single block. On each trial, four photographic images of objects were presented on the left side of a visual display, and four potential destinations were depicted on the right of the screen (see Figure 1). Photographic images were obtained either from the Picture Perfect Stimulus Set (Saryazdi, Bannon, Rodrigues, Klammer, & Chambers, 2018) or were downloaded from free online repositories (materials are available upon request). Each photograph was presented at a 250 × 250 pixel resolution on a 27-in. LCD monitor in a slightly staggered layout (to avoid a reliance on using location descriptions). On each trial, listeners were instructed to move one of these images (located on the left side of the screen) to a specific location within a 2 × 2 grid on the right side of the screen, with each grid quadrant containing a colored geometric shape (purple star, blue heart, green square, and red circle). These same geometric shapes were used on all trials. The placement of both the photographs on the left of the screen and geometric shapes on the right was randomized within these areas across trials. On each critical trial, the target image to be moved was either unrelated to all other photographic images (no-contrast condition), or was accompanied by another object in

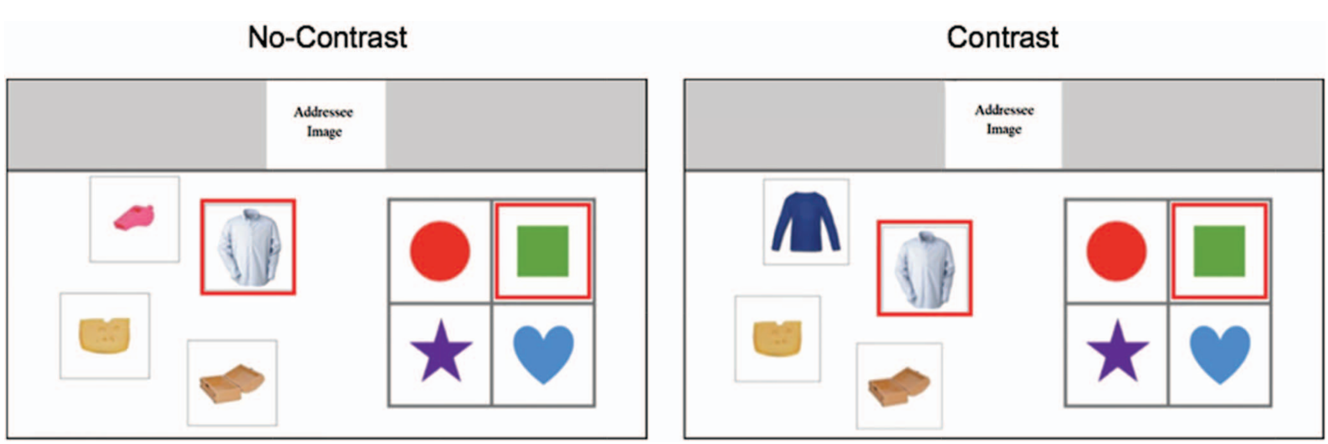

Figure 1. Sample experimental display. See the online article for the color version of this figure.

the same category (contrast condition). The same object served as the target in both conditions. In the contrast condition, a modified basic-level term (e.g., *spotted dog*) or a subordinate term (e.g., *Dalmatian*) would therefore be necessary for referential success (i.e., the unadorned term *dog* would be ambiguous). Note that all target items were selected so they could in principle be differentiated with either a subordinate term or a phrase consisting of a modifier and basic-level noun (see the online supplemental materials for the full list of items). The presence or absence of the contrasting object was varied across lists such that each participant saw a given array of objects in only one experimental condition. However, the pairing of arrays to conditions was cycled across lists such that each object array occurred equally often in both contrast conditions. Filler trials were designed to counteract any strategies or contingencies that might result from the critical trials. For example, on some trials none of the to-be-described objects were from the same category, and on others there were two same-category objects present but neither of these was the item prompted for description.

Procedure

Upon arrival, all participants completed the consent form and language questionnaire. Older adult participants were also screened for hearing, visual acuity, and color blindness. Participants were then told the task involved recording instructions for future participants who would be performing the actions described in these instructions. These future addressees were described as being younger adults under the age of 25, older adults above the age of 65, as well as a computer system capable of speech recognition. Participants were then given two practice trials, followed by an opportunity to ask any clarification questions.

The main task began with an image of a potential addressee and a short description regarding the age of the human addressee or a general statement regarding the computer (e.g., "Next Player: Computer . . . You will now provide instructions to a computer equipped with speech recognition software"). An image of the addressee (i.e., photograph of a younger/older adult performing the "listening" version of the future task on a computer, or a clipart of a microphone for the speech recognition software) remained on the top of the screen across all trials to ensure participants were consistently aware of their intended audience. The image was the same for all trials within a given addressee block. The ordering of the three blocks of trials corresponding to the different addressee types was counterbalanced across participants. The target item and its intended location in each trial were cued in sequence by colored frames such that the first cue (indicating the target) appeared approximately 5 s after the display onset, and was preceded by a short tone to capture participants' attention. The second cue (indicating the destination) appeared 2 s later. Participants were asked to provide their instructions in the form of *Click on the [. . .] and move it to the [. . .].* The recording was captured by high quality directional microphone (Sennheiser MIKE 600) connected to a preamplifier (Eurorack MX 602A), whose output was recorded by computer. The experiment was implemented with Experiment Builder software (SR Research, Ontario, Canada) and lasted approximately 30 min.

Coding Procedures

The recordings obtained from all critical trials were transcribed by a trained coder blinded to experimental condition. Each recording was coded for speech onset latencies (the time from the appearance of the target frame cue to onset of the speaker's utterance), speech rate (syllables per second), and the occurrence of disfluencies, which were coded separately for subcategories such as repairs (e.g., *Click on the alarm clar -cl—alarm clock*) and filled pauses (e.g., *Click on the . . . uhh bed*). Given that the evaluation of unfilled pauses is quite subjective, these were excluded from analyses of disfluency (see Bortfeld et al., 2001; Fox Tree, 1995; Watanabe et al., 2008). Note that the syllable count used in the calculation of speech rate included filled pauses as they contributed to the duration of the utterance. Next, the second author judged the quality of each referring expression in terms of its adequacy (defined in terms of Grice's maxim of quantity) for uniquely identifying the intended referent in relation to the absence or presence of referential contrast. For this measure, the target description could be underspecified (insufficient information for referential identification), just enough (sufficient information), or overspecified (more information than necessary). In addition, each description was coded as to whether the noun was a basic-level term (e.g., dog) or a subordinate term (e.g., Dalmatian). We also coded the incidence of modifier use (whether or not a modifier was included), and the type of modifiers used, when present (color, size, shape, location, with any modifier that did not fit these classifications coded as "other," but in the end, shape modifiers were never used by participants). The method used to count a particular category type accounted for the fact that in some instances, participants used a phrasal modifier to refer to a given property. For example, if participants used a color phrase such as *black-and-white dog* to refer to the target referent, we coded this as containing one modifier. Finally, we coded the position of the modifier in terms of whether it was prenominal (e.g., *yellow-and-black snake*) or postnominal (e.g., *snake on the bottom right*). The coding criteria considered any modifier that, together with the noun, could uniquely identify the target as one modifier. For example, *yellow-and-black snake on the bottom right* would be coded as one prenominal modifier for color, and one postnominal modifier for location. See Table 1 for examples of the coding scheme corresponding to the display depicted in Figure 1.

An independent coder blind to the experimental conditions provided a secondary check of the coding for a random sample of six younger and six older adult participants (25% of overall data). A high degree of interrater agreement was found for all measures (range = 90–100%, for which discrepancies were reconciled) except the speech timing measures which was calculated based on agreement within a 100 ms margin (~75%). For this reason, a second independent coder was asked to recode speech timing measures, and these were complemented with automated Praat analyses. Any remaining discrepancies were checked and corrected by the authors prior to statistical analyses.

Results

All statistical analyses were conducted using R, Version 3.3.3 (R Core Team, 2017). Because we are interested in age-related differences in production of referential expressions, all analyses are based only on the clause containing the target description

Table 1
Example Coding of Descriptions for the Display Depicted in Figure 1

Example phrase	Information level contrast	Information level no-contrast	Modifier type	Modifier position	Noun specificity
Shirt	Underspecified	Just right	n/a	n/a	Basic-level
Dress shirt	Just right	Just right	n/a	n/a	Subordinate
Men's white shirt	Overspecified	Overspecified	Other and color	Prenominal	Basic-level
Shirt on the left-hand side	Just right	Overspecified	Location	Postnominal	Basic-level

Note. n/a = not applicable.

(*Click on the [. . .]*), unless otherwise noted. See Table 2 for descriptives across conditions. We performed mixed-effect analyses using lme4 Version 1.1–15 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest Version 3.0–1 (Kuznetsova, Brockhoff, & Christensen, 2017). The design included age (younger vs. older adults) as a between-participants factor, and contrast condition (contrast vs. no-contrast) and addressee groups (same age, different age, and computer) as within-participant factors. Each model included age (coded as 1 for younger vs. −1 for older adults), contrast condition (coded as 1 for contrast vs. −1 for no-contrast), addressee group, using same-age peers as the reference group (Comparison 1: coded as 1 for same age vs. −1 for different age; Comparison 2: coded as 1 for same age vs. −1 for computer), and their interactions as fixed effects. The same-age peer was selected as a reference group because it was deemed most optimal for a comparison involving both human and computer addressee groups. The tables and figures, however, show the results separately for each addressee group. Unless otherwise noted, the random effect structure in all models was based on a maximal model (Barr, Levy, Scheepers, & Tily, 2013) in which we included random intercept terms for participant and item (object arrays), by-participant slope terms for contrast and addressee group and by-item slope terms for age, contrast, addressee group, as well as all corresponding interaction terms. We report the results of all analyses that reach significance at a .05 alpha level in the main document. Complete statistical summary tables are provided in the online supplemental materials.

Production Performance Measures

Speech onset latencies. This measure captures aspects of linguistic planning in advance of speech output. The results showed a reliable effect of contrast, with longer speech onset latencies in the contrast condition (M = 2.54 s, SD = 1.48) compared to the no-contrast condition (M = 2.34 s, SD = 1.15), β = 0.10, SE = 0.04, $t(15)$ = 2.38, p = .031. Interestingly, however, we did not find any age differences in this measure despite the fact that previous work using isolated picture naming tasks has shown longer speech onset latencies (e.g., Spieler & Griffin, 2006; Mortensen et al., 2008), as well as difficulty with lexical retrieval in older adults (see Mortensen et al., 2006, for a review).[1] We also observed an effect of addressee group, with longer speech onset latencies when speech was directed toward one's own age group (M = 2.44 s, SD = 1.35) than to a computer system (M = 2.36 s, SD = 1.17), β = 0.08, SE = 0.04, $t(585)$ = 2.03, p = .043. However, this difference was not found when speakers were talking to a different age group, β = −0.07, SE = 0.05, $t(25)$ = −1.33, p = .194. It is possible that participants were more concerned with providing adequate instructions to a future human participant than a computer and as such spent more time planning their utterance. A follow-up analysis comparing speech onset latencies toward human versus computer (coded as 1 vs. −1, respectively), showed numerically slower initiation when talking to a human (M = 2.47 s, SD = 1.40) than to a computer addressee (M = 2.36 s, SD = 1.17). However this effect did not reach significance, β = 0.06, SE = 0.03, $t(695)$ = 1.95, p = .051.

Speech rate. The speech rate measure reflects the number of syllables per second in the participant's production of the clause *Click on the [. . .]*. The results revealed no reliable differences across age groups, contrast conditions, or addressee groups.[2] However, we observed an effect approaching significance for the latter manipulation, whereby participants spoke faster to their own age group (M = 3.95, SD = 1.77) than a different age group (M = 3.77, SD = 1.25), β = 0.11, SE = 0.05, $t(54)$ = 2.02, p = .049. A follow-up analysis conducted separately for younger and older adult participants revealed that the effect was driven by younger adults speaking at a slower rate to older adults β = 0.15, SE = 0.06, $t(22)$ = 2.67, p = .014. As previously mentioned, a slow speech rate is one of the common characteristics of "elderspeak" (Kemper, 1994; Kemper et al., 1996, 1998). The observed trend is therefore in line with the previous literature on speech adaptation toward older addressees. What we highlight here is the fact that speech adaptation occurs even in sentences referring to simple, everyday objects, where in theory there should be little reason to assume comprehension difficulties in older listeners. The present study also reveals that such adaptation occurs in the absence of a physical addressee, suggesting the relevant adjustments are made fairly easily.

Disfluency. Overall, the rate of disfluency in the individual subcategories (e.g., repairs, filled pauses) was quite low. We therefore created a composite score including all disfluency types (approximately 10% of all utterances contained a disfluency). This

[1] As an alternative measure, we examined speech onset latency using the beginning of the description following "Click on the" (whether noun or adjective). This analysis did not change the pattern of results observed.

[2] An evaluation of speech rate in the full utterance (including the portion in which speakers described the intended destination for a target object) showed older adults were slower (M = 3.23, SD = 0.96) than younger adults (M = 4.01, SD = 0.92), β= 0.39, SE = 0.11, $t(49)$ = 3.47, p = .001. Recall that the "destination" part of the sentence involved naming the same colored shapes, yet the relative position of these shapes changed on every trial. This may have been more confusing for older adults, leading to the observed pattern.

Table 2
Means and Standard Deviations for Speech Onset Latencies, Speech Rate, Speech Disfluency, and Informational Adequacy Across All Conditions

Measure	No-contrast YA *M* (*SD*)	No-contrast OA *M* (*SD*)	Contrast YA *M* (*SD*)	Contrast OA *M* (*SD*)
Speech onset latencies (s)	**2.57 (1.12)**	**2.11 (1.13)**	**2.75 (1.42)**	**2.32 (1.50)**
Younger addressee	2.56 (1.12)	2.04 (1.01)	2.74 (1.59)	2.48 (1.89)
Older addressee	2.66 (1.17)	2.16 (1.25)	2.84 (1.44)	2.30 (1.35)
Computer	2.48 (1.08)	2.12 (1.14)	2.66 (1.24)	2.17 (1.16)
Speech rate (syllables/s)	**4.02 (0.97)**	**3.80 (1.86)**	**3.97 (1.28)**	**3.72 (1.94)**
Younger addressee	4.05 (0.96)	3.71 (1.57)	4.17 (1.25)	3.66 (1.34)
Older addressee	3.93 (0.89)	3.93 (2.47)	3.77 (1.12)	3.64 (1.98)
Computer	4.09 (1.07)	3.75 (1.38)	3.97 (1.45)	3.87 (2.39)
Speech disfluency (counts)	**0.05 (0.22)**	**0.11 (0.42)**	**0.18 (0.48)**	**0.21 (0.75)**
Younger addressee	0.06 (0.23)	0.15 (0.49)	0.14 (0.48)	0.19 (0.52)
Older addressee	0.06 (0.23)	0.08 (0.28)	0.19 (0.49)	0.30 (1.13)
Computer	0.04 (0.20)	0.10 (0.45)	0.19 (0.46)	0.13 (0.38)
Adequacy (proportion)	**0.92 (0.28)**	**0.79 (0.41)**	**0.80 (0.40)**	**0.63 (0.48)**
Younger addressee	0.90 (0.30)	0.78 (0.42)	0.81 (0.40)	0.56 (0.50)
Older addressee	0.92 (0.28)	0.76 (0.43)	0.82 (0.39)	0.65 (0.48)
Computer	0.93 (0.26)	0.82 (0.39)	0.78 (0.42)	0.70 (0.46)

Note. YA = younger adults; OA = older adults. Bolded values denote averages across addressee groups.

analysis revealed a significant effect of contrast, with more disfluencies occurring when a contrasting same-category object was present (M = 0.19, SD = 0.63) compared to when there was no contrasting object (M = 0.08, SD = 0.33), β = 0.06, SE = 0.02, $t(16)$ = 2.74, p = .014. This is in line with previous findings showing that disfluency increases as a function of task difficulty (e.g., Bortfeld et al., 2001). Nonetheless, we did not observe any differences as a function of age or addressee group.

Measures of Descriptive Content

Informational adequacy. As mentioned earlier we coded referential descriptions in terms of whether they were just enough, overspecified, or underspecified for the purpose of uniquely identifying the intended target within the given context. Figure 2 presents the results for each age group and contrast condition. As the figure illustrates, the rate of underspecified descriptions was quite similar across the two age groups, such that participants apparently failed to consider the contrast object when formulating their description approximately 10% of the time (see also Koolen, Gatt, Goudbeek, & Krahmer, 2011; Gatt et al., 2017). In general, the rate of underspecified descriptions in referential production studies is lower than that for overspecified descriptions unless the experimental paradigm involves a more complex visual context (e.g., Healey & Grossman, 2016). For the statistical analysis, we

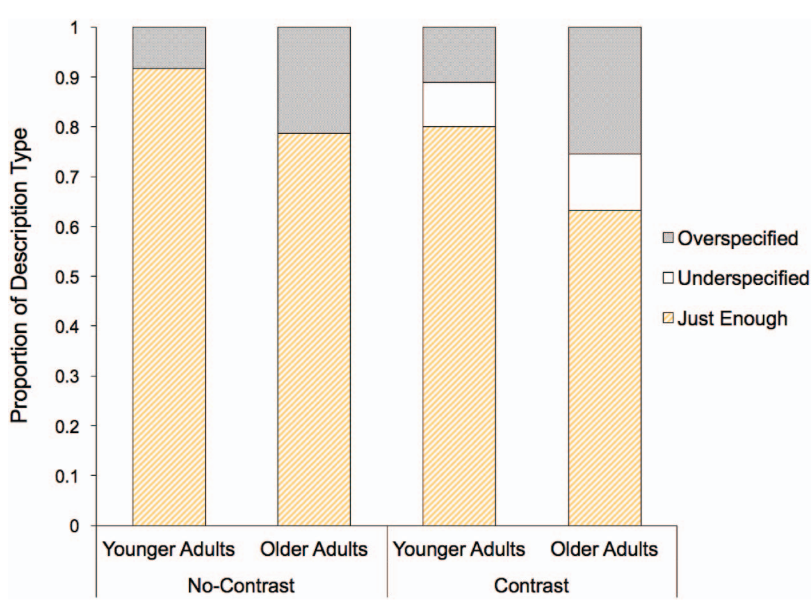

Figure 2. Informational adequacy across conditions. See the online article for the color version of this figure.

used a logistic mixed effect model comparing the incidence of descriptions containing "just enough" information relative to both underspecified descriptions (containing inadequate information for referent identification), and overspecified descriptions (containing a redundant informational unit), with these last two types being grouped together. The results revealed a reliable age-related difference, with younger adults providing more optimal descriptions (M = 0.86, SD = 0.35) in comparison to older adults (M = 0.71, SD = 0.45), β = 0.70, SE = 0.25, Z = 2.84, p = .005. In addition, we observed a significant effect of contrast condition, whereby optimal descriptions were produced more often when referential contrast was absent (M = 0.85, SD = 0.36) than when it was present (M = 0.72, SD = 0.45), β = −0.67, SE = 0.18, Z = −3.64, p < .001. However, we did not observe any Age × Contrast interaction or any effects of audience design.

The coding process revealed that a large number of referring descriptions contained subordinate terms across both contrast conditions, with overall similar rates across the two age groups (see Table 3). In the coding criteria used for the analysis above, an unadorned subordinate term (e.g., *Dalmatian*) in the contrast condition would be coded as "just enough," whereas a subordinate term plus a modifier would be coded as an overspecified description in this condition (e.g., *Dalmatian on the left*). In the no-contrast condition, we coded an unadorned subordinate term as "just enough" because there was no separate linguistic element reflecting redundancy. However, this is not entirely in line with a strict account of Gricean informativity, where the information provided is more than what is necessary (i.e., a basic-level term such as *dog* would suffice). For this reason, and in the interest of evaluating the impact of alternative coding procedures for a task where speakers are free to use subordinate terms, we conducted a secondary analysis whereby subordinate terms in the no-contrast condition were classified as overspecified (yet unadorned subordinates in the contrast condition were coded as "just enough," as in the initial analysis). The alternative coding scheme yielded a similar age effect, with younger adults providing sufficient descriptions more often than older adults. As before, there was also an effect of contrast condition (and no interaction), although the direction of the contrast effect was reversed. Specifically, the data now showed a pattern where optimal descriptions were produced more often when referential contrast was present (M = 0.72, SD = 0.45) compared to when contrast was absent (M = 0.46, SD = 0.50), β = 0.88, SE = 0.28, Z = 3.10, p = .002. This is primarily due to the fact that almost half of the adequate descriptions in the no-contrast condition involved subordinate terms, and the new coding criteria changed the status of these descriptions to overspecified (and hence not optimal). Although the key finding regarding age difference was not affected, these coding-dependent differences highlight the importance of more careful methodological and theoretical treatment of subordinate terms in future studies of referential production—a topic that is often not discussed despite its importance for capturing the referential options speakers face in the real world (although see Yoon & Brown-Schmidt, 2013).

Table 3
Percentage of Descriptions Containing Subordinate Terms by Age, Contrast, and Informational Adequacy

Information level	No-contrast		Contrast	
	YA	OA	YA	OA
Just enough	44%	49%	65%	65%
Overspecified	22%	41%	75%	74%

Note. YA = younger adults; OA = older adults. Remaining percentage reflects the use of basic-level terms.

We next conducted two additional analyses to better understand the observed differences in overspecification across age groups. The first analysis examines the use of modifiers in the referential description and the second is an exploration of the particular type and position of modifiers used.

Modifier use. This analysis was conducted to provide a more direct comparison with past studies that use the presence or absence of a modifier as the primary way to measure the production of successful descriptions in relation to the contrast/no-contrast manipulation. Because the assumptions that underlie this measure do not hold when speakers use subordinate terms, we first restricted the data to only those cases where speakers used basic-level nouns. Trials were then coded using a binary measure capturing whether or not participants produced a nominal modifier. The by-item slope for age, contrast, and addressee interaction was dropped from the model because the full model did not converge. We found a significant effect of contrast condition, β = 21.14, SE = 5.27, Z = 4.01, p < .001, with greater use of modifiers when referential contrast was present (M = 0.75, SD = 0.44) than when it was absent (M = 0.17, SD = 0.38). This clearly demonstrates speakers' awareness of the need to provide more information in the former case. Interestingly, we also found a significant Contrast × Different-Age Addressee interaction, β = 2.79, SE = 1.25, Z = 2.24, p = .025. This pattern arises because speakers (collapsed across age) are more likely to provide modifiers when addressing same-age peers (M = 0.80, SD = 0.40), than addressing different-age peers (M = 0.68, SD = 0.47), but only when referential contrast was present. We also found a significant Age × Contrast × Computer Addressee interaction, β = 3.30, SE = 1.25, Z = 2.65, p = .008. As illustrated in Figure 3, this reflects a pattern whereby older adults are more likely to overspecify the referent in the no-contrast condition when speaking to a same-age addressee (M = 0.32, SD = 0.47) than when speaking to a computer (M = 0.12, SD = 0.33).

Modifier type and position. Our final set of analyses explore the type of property expressed in the modifier as well as the modifier's syntactic expression. Figure 4 shows the relative proportion of modifier by property type, across age groups and contrast conditions. Color was the dominant type of modifier produced by both age groups, consistent with previous studies of referential production (e.g., Belke & Meyer, 2002; Fukumura, 2018; Tarenskeen et al., 2015), which typically note a relationship between frequently expressed properties and their perceptual salience. Interestingly, older adults were also more likely to use "other" modifiers such as gender (e.g., *the men's shirt*) or material (e.g., *plastic ball*)—subtypes that were never used by the younger adults (see the online supplemental materials for more examples). Also, whereas older adults showed a tendency to use a broader range of modifier types beyond the typical types (color, size, location) mentioned above in the contrast condition, they did less so in the no-contrast condition (44% vs. 20%, respectively). Indeed, older

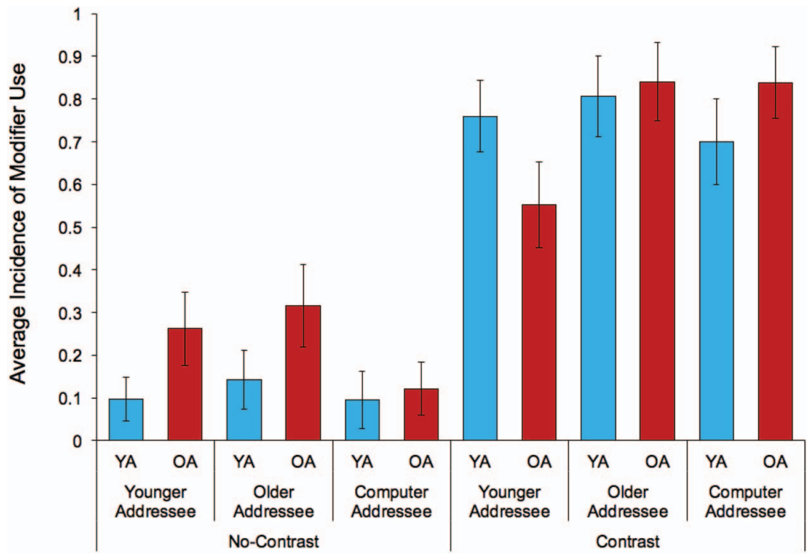

Figure 3. Average incidence of modifier use by younger adults (YA) and older adults (OA) across contrast and addressee conditions (data is limited to descriptions containing basic-level nouns). See the online article for the color version of this figure.

adults' rate of using uncommon modifiers was equivalent to their rate of using color modifiers in the contrast condition. To evaluate this pattern, we calculated a measure reflecting the rate with which each participant used "other" modifiers. The analysis was based on an intercept-only model (full model did not converge) and revealed a main effect of age, $\beta = -0.30$, $SE = 0.10$, $t(46) = -3.07$, $p = .004$, a main effect of contrast condition, $\beta = 0.66$, $SE = 0.09$, $t(46) = 7.12$, $p < .001$, and an Age $\times$ Contrast interaction, $\beta = -0.22$, $SE = 0.09$, $t(46) = -2.37$, $p = .022$. Further analyses confirmed that, in the contrast condition, the use of "other" modifiers by older adults ($M = 2.17$, $SD = 1.52$) was greater than that of younger adults ($M = 1.12$, $SD = 0.8$), $t(35) = 2.97$, $p = .005$. No analogous effect was found in the no-contrast condition ($p = .271$). This condition instead showed a pattern where older adults increased their use of location modifiers, which is notable because this type was entirely absent in younger adults' descriptions in this condition.

Although older adults' use of uncommon descriptions has been shown in referential communication tasks with comparatively complex stimuli (e.g., Hupet et al., 1993), the current results suggest that such patterns of communication persist even for everyday familiar objects. Because it is possible that greater variability in the use of modifiers might be related to the increased vocabulary knowledge that is characteristic of older adults, we conducted a correlation analysis within the older adult sample relating the use of "other" modifiers with Mill Hill vocabulary scores. The results revealed no significant relationship between the two, $r(22) = .12$, $p = .582$. We next explored whether unnecessary modifiers (used with basic-level nouns) tended to occur before or after the noun, to understand whether prenominal modifiers might be used to buy time in lexical selection for the noun term. Overall, both age groups were more likely to use prenominal modifiers such as *white shirt* than postnominal ones such as *the shirt on the right-hand side*. However, older adults showed a greater tendency to use postnominal modifiers in their overspecified descriptions (29% postnominal only, 56% prenominal only, 15% both), with very few being produced by younger adults (0% postnominal only, 85% prenominal only, 15% both). An analysis of the incidence of postnominal versus prenominal redundant modifiers across age groups revealed a significant effect ($p = .005$, using Fisher's exact test due to the low number of observations meeting the criteria). We then examined the occurrence of postnominal modifiers in contexts where modifiers were necessary (i.e., when basic-level nouns were used in the condition with referential contrast). In these cases (18 in total), we saw no age-related differences in the descriptions produced by younger adults (eight) and older adults (10).

Discussion

Effective communication depends on both the clarity of spoken language and the adequacy of the information provided by interlocutors. Older adults are often portrayed as less efficient communicators, but much of the primary evidence for this idea is drawn from studies using comparatively complex tasks that may place greater demand on older adults' cognitive processes (e.g., Bortfeld et al., 2001; Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Thus, it is quite possible that reported declines in language performance could be induced in part by the nature of the experiment itself. In the present study, we explored age-related differences in language production using a task where speakers provided spoken instructions relating to color photographs of everyday objects. Younger and older participants named objects for an imagined addressee, occurring in either contrastive contexts (where the presence of another object in the same category necessitated the use of a more elaborate description) or noncontrastive contexts. We also examined whether participants' referential expressions varied as a function of addressee type, namely when

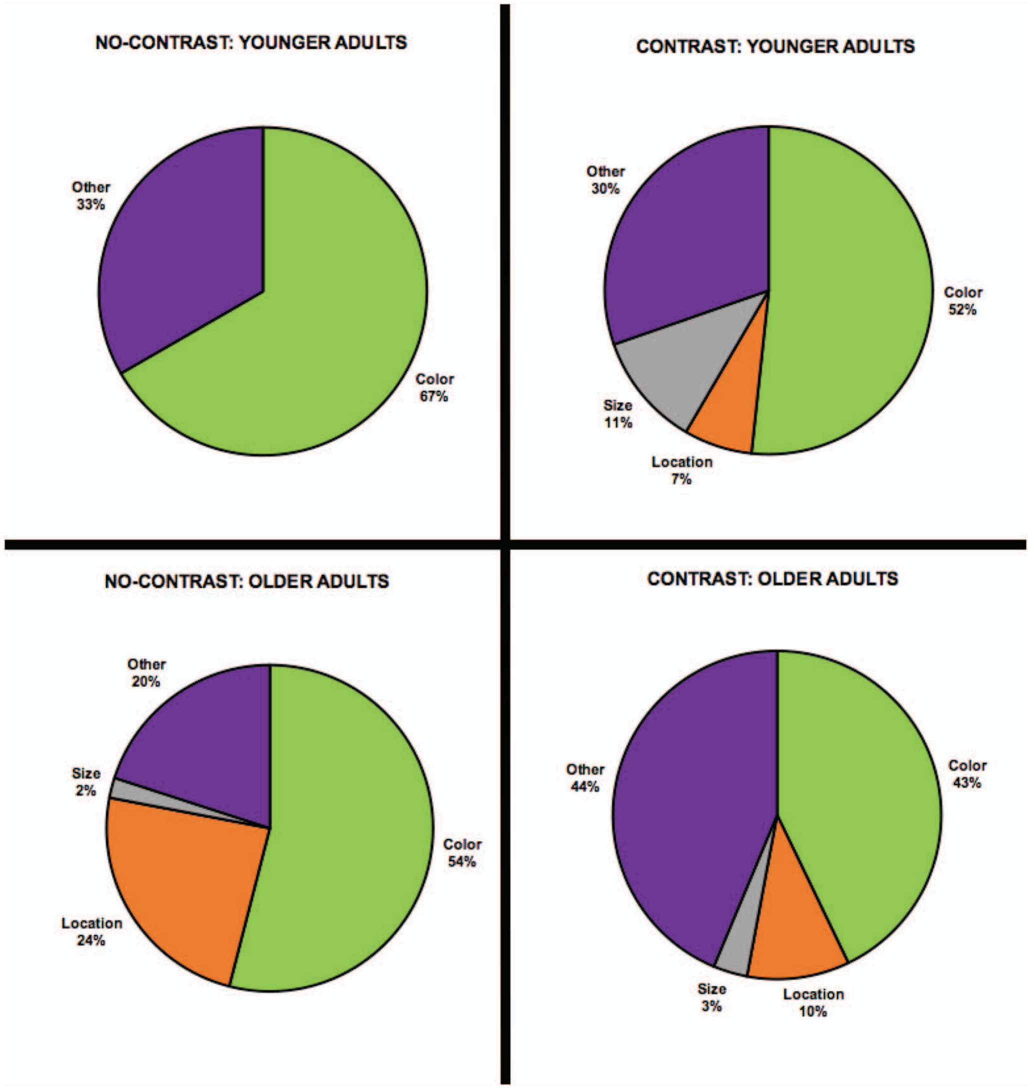

Figure 4. Breakdown of modifier types by age and contrast conditions. See the online article for the color version of this figure.

speaking to an imagined younger adult, older adult, or an automated dialogue system. A variety of measures including speech onset latencies, speech rate, fluency, and the informativity of speakers' descriptions were included to provide a comprehensive picture of communicative ability within a single paradigm.

The measure of speech onset latencies (understood to reflect utterance planning processes), revealed that speakers were slower when target objects occurred alongside another same-category object, reflecting the additional planning costs required for referential production in this context[3]. However, older adults were not overall slower than younger adults, nor was there an interaction between age and context type. Because our visual stimuli allowed for the use of either basic-level or subordinate terms (potentially increasing lexical competition), any differences in relation to past studies are unlikely to be due to linguistic aspects of our materials (i.e., a more restricted range of options for naming). Another possible explanation is that the preview interval may have provided participants with ample time to formulate referential descriptions, reducing the potential to detect age-related effects. This interval, however, is not unrepresentative of real-world contexts, where candidate referents are a reasonably stable element of the situational context, and where individuals determine their own conversational pacing. Further, an effect of contrast condition on speech onset latencies was nonetheless observed, suggesting our measure was sensitive enough to capture relevant differences. Future research is clearly necessary to determine the extent to which age differences in speech onset latencies might arise from task demands. In terms of addressee type, we found slower speech

[3] To examine whether the use of subordinate versus basic-level terms was related to the amount of time spent inspecting the scene before speaking, we calculated average speech onset latencies for each noun type. Latencies for basic-level and the subordinate terms were in fact similar (basic term = 2.53 s; subordinate = 2.35 s).

onset times occurring when speakers addressed a same-age peer than a computer. We speculate that perhaps speakers are more concerned with a peer's comprehension than a computer's, and thus take longer to plan their utterance.

In the speech rate measure, we did not observe any main effect of age or an effect of contrast condition. The lack of an age effect for this measure is consistent with some past studies. For example, analyses of natural phone conversations have revealed only a small difference between younger and older adults' speech rate (Horton et al., 2010). We did, however, observe an effect of audience design on speech rate, with younger adults speaking more slowly with an older addressee. This slower speech rate is one of the common characteristics of "elderspeak" that younger speakers will often adopt when talking to older adults (Kemper, 1994; Kemper et al., 1996), and the current results extend existing evidence for this type of accommodation in intergenerational communication. More specifically, the results show that younger adults will adapt speech directed toward older adults even when the addressee is imagined rather than physically present. Older adults, in contrast, did not show this form of adaptation, which has also been reported in past work (Horton & Spieler, 2007; Kemper & Harden, 1999; Kemper et al., 1996).

In our speech fluency measure, both younger and older adults were significantly more disfluent in the contrast condition, consistent with the idea that greater processing demands will lead to an increased incidence of disfluencies (e.g., Corley & Stewart, 2008). But once again, we did not observe age-related differences, in line with some earlier studies (e.g., Duchin & Mysak, 1987; Griffin & Spieler, 2006). Fluency rates were also not affected by the audience type manipulation. Overall, the absence of age effects in our timing and fluency measures may be related to the simple everyday objects we used as stimuli, substantiating the idea that task complexity is an important consideration in studies of language production and aging. It is also possible that the absence of age effects in the present study results from having a fairly homogeneous and high-functioning sample of older adults. Separate assessments of cognitive abilities (e.g., memory, inhibition, processing speed) in future studies would help clarify the role of individual differences at this level.

The present study did, however, find reliable age-related differences involving the informativity and content of speaker's referential descriptions. Specifically, older adults were less optimally Gricean in their descriptions in comparison to younger adults. This finding is broadly consistent with Healey and Grossman (2016), although there are important differences. For example, whereas those authors reported a greater incidence of both overspecification and underspecification, the current study only showed age-related differences in terms of overspecification. This may be because Healey and Grossman varied the presence of zero, one, or three same-category contrasting items within a perspective-taking paradigm, thus increasing the complexity of the task. Indeed, they found that the older adults' likelihood of producing underspecified descriptions increased when there was a greater number of contrasting objects (requiring multiple modifiers for referential success). In comparison, the present study only included a single same-category object in the contrast condition, which may in part explain why we did not find similar patterns in the rate of underspecification. Another feature of the current study was the potential to use subordinate terms to refer to target objects, which may help ensure a smaller overall incidence of underspecification. Subordinate terms were in fact used very often, and were produced by older and younger adults at a similar rate. Nonetheless, as in Healey and Grossman, the present results did show greater overspecification by older adults relative to younger adults, particularly in the no-contrast condition where modification was unnecessary. Further, in addition to the differences observed in rates of overspecification, we also found differences in the kinds of modifiers used by younger and older adults. Specifically, older adults used an overall broader range of modifiers, including some that were never observed in the speech of our younger participants.

What might underlie the differences found in older adults' referential production? For the increased tendency to overspecify referents, there are several possible explanations. One is that this behavior might be understood as a special instance of older adults' tendency for so-called off-topic verbosity, although this seems unlikely because the behavior is evident primarily when producing autobiographical descriptions or describing complex rather than simple visual stimuli (Arbuckle, Nohara-LeClair, & Pushkar, 2000; James, Burke, Austin, & Hulme, 1998; Pushkar et al., 2000; Yin & Peng, 2016). Given older adults' purported difficulties with lexical retrieval, another possibility is that overspecification by means of prenominal modifiers (which arguably denote properties that are easy to extract from the visual scene) might allow an older adult to buy time while still trying to identify a noun for the target. However, about a third of superfluous modifiers produced by older adults occurred only in the postnominal position (none of the overspecified descriptions produced by younger adults contained postnominal modifiers alone), which would be unexpected if the modifier was added simply to buy additional time to retrieve the target noun. Further, on this account, we would also expect age differences in the other measures such as speech onset latencies and disfluency, which were not observed. Yet another explanation is that speakers choose to overspecify as a way to reduce cognitive effort (namely, the need to securely establish that the target is the only instance of its category). In other words, overspecification would help guarantee identification of the intended object (Davies & Arnold, 2018). The participants in the current task were also asked to provide instructions in a game-like paradigm so a future addressee could identify the target object. Perhaps older adults take this goal of the communication game more seriously, and/or overspecify as a strategy for ensuring referential success (Arts, Maes, Noordman, & Jansen, 2011).

As for older adults' greater use of uncommon modifier types, our results are consistent with previous evidence that older adults tend to exhibit greater lexical diversity, namely the use of terms that are otherwise rare in the context of the study (Dennis & Hess, 2016; Hupet et al., 1993; Horton et al., 2010; Kavé, Samuel-Enoch, & Adiv, 2009). Could this pattern be explained in terms of older adults' greater vocabulary knowledge? One important observation is that the difference in question involves comparatively simple words (e.g., *men's shirt* vs. *white shirt*), and as such is difficult to relate to the notion of expanded vocabulary. Further, older adults' Mill Hill vocabulary scores showed no relationship with their individual tendency to use uncommon modifiers. A fully distinct explanation is that the tendency to use uncommon modifiers could be an age cohort-related effect, reflecting generational differences in communication style. One question raised by such a possibility is whether older adults process descriptions with uncommon and/or redundant modifiers differently than younger adults in the course of real-time spoken language comprehension. Research on this question could clarify the strength of the connection between language production and comprehension systems across the adult lifespan.

Conclusion

In the present study, we observed age-related differences in language production between younger and older adults in a context involving reference to everyday objects. However, not all aspects of production were influenced to the same extent. The most striking differences involved older adults' increased tendency to provide "too much" information in referential descriptions, along with greater variety in lexical choices. In contrast, measures of speaking performance (speech onset latencies, speech rate, and fluency) were mostly stable across age groups. Importantly, older adults' production patterns had little or no impact on the *success* of their referential descriptions. We did not, for example, find that older adults had an overall greater tendency to underspecify their descriptions, which would lead to referential failure. Similarly, the use of less-common modifiers might be understood as a deliberate strategy to increase the success of conveying communicative intentions. Thus, the observed age-related changes in production still meet the need of providing relevant (even if superfluous) information for effective communication. Taken together, the results not only inform our understanding of patterns of change in referential production across the adult lifespan, but also highlight the importance of employing paradigms that are representative of everyday communication.

References

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies, 38,* 547–566. http://dx.doi.org/10.1006/imms.1993.1026

Arbuckle, T. Y., Nohara-LeClair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging, 15,* 65–77. http://dx.doi.org/10.1037/0882-7974.15.1.65

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43,* 361–374. http://dx.doi.org/10.1016/j.pragma.2010.07.013

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,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

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

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional object discrimination: Analyses of viewing patterns and viewing times during "same"–"different" decisions. *The European Journal of Cognitive Psychology, 14,* 237–266. http://dx.doi.org/10.1080/09541440143000050

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes, 22,* 1178–1211. http://dx.doi.org/10.1080/01690960701461541

Bell, L., Gustafson, J., & Heldner, M. (2003, August). Prosodic adaptation in human-computer interaction. *Proceedings of ICPHS, 3,* 833–836.

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). "Older is always better": Age-related differences in vocabulary scores across 16 years. *Psychology and Aging, 30,* 856–862. http://dx.doi.org/10.1037/pag0000051

Bergmann, K., Branigan, H. P., & Kopp, S. (2015). Exploring the alignment space–lexical and gestural alignment with real and virtual humans. *Frontiers in ICT, 2,* 7. http://dx.doi.org/10.3389/fict.2015.00007

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44,* 123–147. http://dx.doi.org/10.1177/00238309010440020101

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition, 121,* 41–57. http://dx.doi.org/10.1016/j.cognition.2011.05.011

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1482–1493. http://dx.doi.org/10.1037/0278-7393.22.6.1482

Brown-Schmidt, S., & Tanenhaus, M. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54,* 592–609. http://dx.doi.org/10.1016/j.jml.2005.12.008

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science, 13,* 21–24. http://dx.doi.org/10.1111/j.0963-7214.2004.01301006.x

Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 373–443). New York, NY: Psychology Press.

Campbell, K. L., Samu, D., Davis, S. W., Geerligs, L., Mustafa, A., Tyler, L. K., . . . the for Cambridge Centre for Aging and Neuroscience. (2016). Robust resilience of the frontotemporal syntax system to aging. *The Journal of Neuroscience, 36,* 5214–5227. http://dx.doi.org/10.1523/JNEUROSCI.4561-15.2016

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland Publishing Co.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201–242. http://dx.doi.org/10.1006/cogp.1998.0693

Corley, M., & Stewart, O. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass, 2,* 589–602. http://dx.doi.org/10.1111/j.1749-818X.2008.00068.x

Coupland, N., & Giles, H. (1988). Introduction: The communicative contexts of accommodation. *Language & Communication, 8,* 175–182. http://dx.doi.org/10.1016/0271-5309(88)90015-8

Dahlgren, D. J. (1998). Impact of knowledge and age on tip-of-the-tongue rates. *Experimental Aging Research, 24,* 139–153. http://dx.doi.org/10.1080/036107398244283

Davies, C., & Arnold, J. E. (2018). Reference and informativeness: How context shapes referential choice. *Handbook of Experimental Semantics and Pragmatics*. Oxford, United Kingdom: Oxford University Press.

Davies, C., & Katsos, N. (2013). Are speakers and listeners "only moderately Gricean"? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics, 49,* 78–106. http://dx.doi.org/10.1016/j.pragma.2013.01.004

Davies, C., & Kreysa, H. (2017). Looking at a contrast object before speaking boosts referential informativeness, but is not essential. *Acta Psychologica, 178,* 87–99. http://dx.doi.org/10.1016/j.actpsy.2017.06.001

Dennis, P. A., & Hess, T. M. (2016). Aging-related gains and losses associated with word production in connected speech. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 23,* 638–650. http://dx.doi.org/10.1080/13825585.2016.1158233

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders, 20,* 245–257. http://dx.doi.org/10.1016/0021-9924(87)90022-0

Elsner, M., Clarke, A., & Rohde, H. (2018). Visual complexity and its effects on referring expression generation. *Cognitive Science, 42,* 940–973. http://dx.doi.org/10.1111/cogs.12507

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54,* 554–573. http://dx.doi.org/10.1016/j.jml.2005.12.009

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104–113. http://dx.doi.org/10.1037/0012-1649.20.1.104

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34,* 709–738. http://dx.doi.org/10.1006/jmla.1995.1032

Fraundorf, S. H., & Watson, D. G. (2013). Alice's adventures in *um*-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience, 29,* 1083–1096. http://dx.doi.org/10.1080/01690965.2013.832785

Fukumura, K. (2018). Ordering adjectives in referential communication. *Journal of Memory and Language, 101,* 37–50. http://dx.doi.org/10.1016/j.jml.2018.03.003

Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science, 41,* 1457–1492. http://dx.doi.org/10.1111/cogs.12375

Giles, H., & Gasiorek, J. (2011). Intergenerational communication practices. In K. W. Schaie & S. L. Willis (Eds.), *The handbooks of aging consisting of three Vols. Handbook of the psychology of aging* (pp. 233–247). San Diego: Elsevier Academic Press. http://dx.doi.org/10.1016/B978-0-12-380882-0.00015-2

Gould, O. N., & Shaleen, L. (1999). Collaboration with diverse partners: How older women adapt their speech. *Journal of Language and Social Psychology, 18,* 395–418. http://dx.doi.org/10.1177/0261927X99018004003

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3, speech acts* (pp. 41–58). New York, NY: Academic Press.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10,* 603–609. http://dx.doi.org/10.3758/BF03196521

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279. http://dx.doi.org/10.1111/1467-9280.00255

Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements. *Brain and Language, 99,* 272–288. http://dx.doi.org/10.1016/j.bandl.2005.08.003

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine, 29,* 737–752. http://dx.doi.org/10.1016/j.cger.2013.07.002

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. http://dx.doi.org/10.1177/0956797614567339

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology, 57,* 169–177. http://dx.doi.org/10.1027/1618-3169/a000021

Healey, M. L., & Grossman, M. (2016). Social coordination in older adulthood: A dual-process model. *Experimental Aging Research, 42,* 112–117. http://dx.doi.org/10.1080/0361073X.2015.1108691

Heller, D., & Chambers, C. G. (2014). Would a blue kite by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language, 70,* 53–67. http://dx.doi.org/10.1016/j.jml.2013.09.008

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47,* 589–606. http://dx.doi.org/10.1016/S0749-596X(02)00019-0

Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22,* 281–290. http://dx.doi.org/10.1037/0882-7974.22.2.281

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging, 25,* 708–713. http://dx.doi.org/10.1037/a0019424

Hummert, M. L., Shaner, J. L., Garstka, T. A., & Henry, C. (1998). Communication with older adults: The influence of age stereotypes, context, and communicator age. *Human Communication Research, 25,* 124–151. http://dx.doi.org/10.1111/j.1468-2958.1998.tb00439.x

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8,* 339–346. http://dx.doi.org/10.1037/0882-7974.8.3.339

James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of "verbosity" in younger and older adults. *Psychology and Aging, 13,* 355–367. http://dx.doi.org/10.1037/0882-7974.13.3.355

James, L. E., Chambers, B. N., & Placzek, C. L. (2018). How scenes containing visual errors affect speech fluency in young and older adults. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 25,* 520–534. http://dx.doi.org/10.1080/13825585.2017.1337061

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 24,* 508–527. http://dx.doi.org/10.1080/13825585.2016.1226249

Kavé, G., Samuel-Enoch, K., & Adiv, S. (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging, 24,* 17–27. http://dx.doi.org/10.1037/a0014579

Kemper, S. (1994). Elderspeak: Speech accommodations to older adults. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 1,* 17–28. http://dx.doi.org/10.1080/09289919408251447

Kemper, S., Ferrell, P., Harden, T., Finter-Urczyk, A., & Billington, C. (1998). Use of elderspeak by young and older adults to impaired and unimpaired listeners. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 5,* 43–55. http://dx.doi.org/10.1076/anec.5.1.43.22

Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging, 14,* 656–670. http://dx.doi.org/10.1037/0882-7974.14.4.656

Kemper, S., Othick, M., Warren, J., Gubarchuk, J., & Gerhing, H. (1996). Facilitating older adults' performance on a referential communication task through speech accommodations. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition, 3,* 37–55. http://dx.doi.org/10.1080/13825589608256611

Kemper, S., Vandepute, D., Rice, K., Cheung, H., & Gubarchuk, J. (1995). Speech adjustments to aging during a referential communication task. *Journal of Language and Social Psychology, 14,* 40–59. http://dx.doi.org/10.1177/0261927X95141003

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics, 43,* 3231–3250. http://dx.doi.org/10.1016/j.pragma.2011.06.008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82,* 1–26. http://dx.doi.org/10.18637/jss.v082.i13

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

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. http://dx.doi.org/10.1016/j.cognition.2017.09.004

Lysander, K., & Horton, W. S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes, 49,* 29–60. http://dx.doi.org/10.1080/0163853X.2011.625547

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age related effects on speech production: A review. *Language and Cognitive Processes, 21,* 238–290. http://dx.doi.org/10.1080/01690960444000278

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2008). Speech planning during multiple-object naming: Effects of ageing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 61,* 1217–1238. http://dx.doi.org/10.1080/17470210701467912

Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A., Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition, 111,* 46–54. http://dx.doi.org/10.1016/j.cognition.2008.12.004

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech & Language, 9,* 19–36. http://dx.doi.org/10.1006/csla.1995.0002

Oviatt, S., Bernard, J., & Levow, G. A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech, 41,* 419–442. http://dx.doi.org/10.1177/002383099804100409

Oviatt, S., MacEachern, M., & Levow, G. A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication, 24,* 87–110. http://dx.doi.org/10.1016/S0167-6393(98)00005-3

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27,* 89–110. http://dx.doi.org/10.1515/ling.1989.27.1.89

Pushkar, D., Basevitz, P., Arbuckle, T., Nohara-LeClair, M., Lapidus, S., & Peled, M. (2000). Social behavior and off-target verbosity in elderly people. *Psychology and Aging, 15,* 361–374. http://dx.doi.org/10.1037/0882-7974.15.2.361

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders, 16,* 217–226. http://dx.doi.org/10.1016/0021-9924(83)90035-7

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6,* 5–42. http://dx.doi.org/10.1111/tops.12078

Raven, J. C. (1965). *Guide to using the Mill Hill vocabulary test with progressive matrices*. London, United Kingdom: H. K. Lewis.

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.3.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. http://dx.doi.org/10.1016/0010-0285(76)90013-X

Rubio-Fernández, P. (2016). How redundant are redundant colour adjectives? An efficiency based analysis of color overspecification. *Frontiers in Psychology, 7,* 153. http://dx.doi.org/10.3389/fpsyg.2016.00153

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by Mechanical Turk and laboratory participants. *Behavior Research Methods, 50,* 2498–2510. http://dx.doi.org/10.3758/s13428-018-1028-5

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology, 60,* 362–367. http://dx.doi.org/10.1037/0022-3514.60.3.362

Schmader, C., & Horton, W. S. (2019). Conceptual effects of audience design in human–computer and human–human dialogue. *Discourse Processes, 56,* 170–190.

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,* 254–271.

Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P., & Tyler, L. K. (2007). On the tip-of-the-tongue: Neural correlates of increased word-finding failures in normal aging. *Journal of Cognitive Neuroscience, 19,* 2060–2070. http://dx.doi.org/10.1162/jocn.2007.19.12.2060

Shafto, M. A., James, L. E., Abrams, L., Tyler, L. K., & Cam-CAN. (2017). Age-related increases in verbal knowledge are not associated with word finding problems in the Cam-CAN cohort: What you know won't hurt you. *The Journals of Gerontology: Series B, 72,* 100–106. http://dx.doi.org/10.1093/geronb/gbw074

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

Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research, 30,* 522–529. http://dx.doi.org/10.1044/jshr.3004.522

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21,* 291–321. http://dx.doi.org/10.1080/01690960400002133

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50,* 163–178. http://dx.doi.org/10.1016/j.specom.2007.07.005

Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology, 6,* 1703. http://dx.doi.org/10.3389/fpsyg.2015.01703

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication, 49,* 2–7. http://dx.doi.org/10.1016/j.specom.2006.10.003

van Lierop, K., Goudbeek, M., & Krahmer, E. (2012). Conceptual alignment in reference with artificial and human dialogue partners. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1066–1071). Austin, TX: Cognitive Science Society.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech, 47,* 297–308. http://dx.doi.org/10.1177/00238309040470030401

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50,* 81–94. http://dx.doi.org/10.1016/j.specom.2007.06.002

Yin, S., & Peng, H. (2016). The role of inhibition in age-related off-topic verbosity: Not access but deletion and restraint functions. *Frontiers in Psychology, 7,* 544. http://dx.doi.org/10.3389/fpsyg.2016.00544

Yoon, S. O., & Brown-Schmidt, S. (2013). Lexical differentiation in language production and comprehension. *Journal of Memory and Language, 69,* 397–416. http://dx.doi.org/10.1016/j.jml.2013.05.005

Yoon, S. O., & Stine-Morrow, E. A. L. (2019). Evidence of preserved audience design with aging in interactive conversation. *Psychology and Aging*. Advance online publication. http://dx.doi.org/10.1037/pag0000341

Received November 7, 2018
Revision received May 7, 2019
Accepted May 11, 2019 ■