Routledge
Taylor & Francis Group

Aging, Neuropsychology, and Cognition
A Journal on Normal and Dysfunctional Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/nanc20

Disfluency across the lifespan: an individual differences investigation

Paul E. Engelhardt & Ioanna Markostamou

To cite this article: Paul E. Engelhardt & Ioanna Markostamou (20 May 2024): Disfluency across the lifespan: an individual differences investigation, Aging, Neuropsychology, and Cognition, DOI: 10.1080/13825585.2024.2354958

To link to this article: https://doi.org/10.1080/13825585.2024.2354958

AGING, NEUROPSYCHOLOGY, AND COGNITION
https://doi.org/10.1080/13825585.2024.2354958

Disfluency across the lifespan: an individual differences investigation

Paul E. Engelhardt[a] and Ioanna Markostamou[b]

[a]School of Psychology, University of East Anglia, Norwich, UK; [b]Department of Psychology, Bournemouth University, Poole, UK

ABSTRACT
This study had two research objectives. The first was to examine age-related differences in the fluency of speech outputs, as prior research contains conflicting findings concerning whether older adults produce more disfluency than younger adults. The second was to examine cognitive individual differences, and their relationship with the production of disfluency. One hundred and fifty-four adults completed a story re-telling task, and a battery of cognitive measures. Results showed that younger adults produced more *um's* and fewer repetitions. For individual differences, results showed that inhibition and set shifting were related to the production of repetitions, and inhibition and working memory were related to uh production. Our results provide clarification about mixed findings with respect age and disfluency production. The individual differences provide clarification on theoretical arguments for disfluent speech in aging (e.g. *Inhibition Deficit Hypothesis*), and also sheds light on the role of executive functions in models of language production.

ARTICLE HISTORY
Received 22 August 2023
Accepted 8 May 2024

KEYWORDS
Aging; speech fluency; disfluency; executive function; individual differences

The first goal of this study was to investigate age-related differences in disfluency production. The literature on speech disfluencies across the lifespan is relatively sparse and filled with conflicting findings. The second goal of this study was to investigate the production of disfluency in aging with respect to cognitive individual difference variables, including executive functions (i.e., working memory, inhibition, and set shifting) and intelligence (i.e., matrix reasoning and vocabulary). Several recent studies have documented significant relationships between different forms of disfluency and cognitive individual differences in younger adults (e.g., Engelhardt et al., 2010). Thus, the second goal of the study was to shed light on how changes in cognitive abilities (e.g., executive functioning), which tend to decrease over the course of aging, may impact the production of disfluency. If cognitive individual differences are related to disfluency production, then we expect increased rates of disfluency in older adults. However, the current study used a cross-sectional design, and thus, cannot definitely establish a causal connection

CONTACT Paul E. Engelhardt p.engelhardt@uea.ac.uk School of Psychology, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13825585.2024.2354958.

between declines in cognitive individual differences and increased disfluency production (in older adults).

In the remainder of the Introduction, we first broadly cover age-related changes in language, and the mixed findings with respect to disfluency production in aging. Second, we review literature on about the relationships between cognitive individual differences and disfluency production. Finally, we outline the rationale and predictions for the current study, with a particular emphasis on what results mean for theories of aging (e.g., *the Inhibition Deficit Hypothesis*).

Age-related language changes

Many studies have shown a range of different language problems across the lifespan with respect to language outputs (for reviews, see Burke & Shafto, 2004, 2008). In general, older adults produce more off topic speech, which is believed to be related to difficulty inhibiting irrelevant information (James et al., 1998). Some studies indicate that the syntactic structures of older adults are less complex (i.e., lower mean number of clauses per utterance), and these differences have been linked to working memory limitations (Kemper et al., 1989; Kemper & Rash, 1988; cf.; Gordon et al., 2019). In picture-naming tasks, older adults (>70 years of age) often produce more errors (i.e., produce semantically-related words) and non-normative (i.e., acceptable but atypical) responses (LaGrone & Spieler, 2006; Verhaegen & Poncelet, 2013). In addition, while several studies have found object naming accuracy to be well-preserved in older age (Goulet et al., 1994; Markostamou & Coventry, 2022; Wierenga et al., 2008), naming speed diminishes (Goulet et al., 1994; Mortensen et al., 2008). In spoken discourse, older adults also have a tendency to produce a greater number of pronouns, indefinite references, and ambiguous expressions (Kemper, 1992; Schmitter-Edgecombe et al., 2000).

As mentioned above, one common language problem reported by older adults is word-finding difficulties (Burke et al., 1991; Gordon & Kindred, 2011; Heine et al., 1999; Schmitter-Edgecombe et al., 2000). These word retrieval problems are linked to disfluency (i.e., higher rates of pauses, interruptions, and non-normative word use). Some researchers have speculated that increased use of non-normative words is a compensatory mechanism that helps older adults deal with word retrieval difficulties (Dennis, 2012). Some associate these difficulties with deficits in inhibitory control (Hasher, 2015). This is because the production system cannot inhibit competing responses, which blocks access to a target word. In contrast, others have suggested that word retrieval difficulties are due to a specific information-transmission deficit (James & Burke, 2000), which affects lexical-to-phonological representations within the production system (Bock & Levelt, 1994; Dell, 1986; Dell & O'Seaghdha, 1992; Dell et al., 1999; Levelt, 1989). Thus, older adults more commonly experience word retrieval difficulties, and these difficulties result in several issues, including potential increased rates of disfluency. In summary, all of these findings are indicative of some language functions becoming less efficient with increasing age, but at the same time, other language functions remain relatively intact (Shafto & Tyler, 2014; Wierenga et al., 2008).

In our review of published English-language articles, we identified 14 studies that examined age-related changes in disfluency production, that is, comparisons of disfluency rates between younger and older adults (Albert, 1980; Bortfeld et al., 2001; Castro &

Table 1. Summary of existing literature on disfluency production in aging, effect sizes in parentheses where available.

	Sample Size							
	YA	OA	Total Disfluent	Filled	Unfilled	Repetitions	Repairs	Task
Albert (1980)	*X*		*X*			*X*	*X*	*Dialogue*
Bortfeld et al. (2001)	*16*	*16*	*X*	*X*[2]		*NS*	*NS*	*Dialogue*
Castro and James (2014)	*29*	*29*	*X (.10)*			*NS*	*NS*	*Picture Description*
Cooper (1990)	*(80 total)*[6]			*X*[4] *(.02)*		*X (.03)*	*NS*	*Picture Description*
Davidson et al. (2003)	*48*	*48*	*NS*					*Sentence Production*
Duchin and Mysak (1987)	*(75 total)*[6]		*NS*			*NS*	*NS*	*Varied*
Horton et al. (2010)	*(300 total)*[6]			*X (.17)*[1]				*Dialogue*
James et al. (2018)	*20*	*20*	*X (.13)*[2]			*X (.23)*	*X (.22)*	*Picture Description*
Kemper et al. (2003)*(E1)*	*30*	*30*	*X (.11)*					*Sentence Production*
Kemper et al. (2009)	*40*	*40*		*X*[5]				*Monologue*
Martin (2017)	*(359 total)*[6]		*X*[3]					*Dialogue*
Schmitter-Edgecombe et al. (2000)	*26+26*[7]	*26*		*NS (.00)*	*NS (.01)*	*NS (.01)*	*X (.11)*	*Varied*
Shewan & Henderson, (1988)	*(60 total)*[6]					*NS*		*Dialogue*
Spieler and Griffin (2006)	*17*	*17*	*X*					*Sentence Production*
Summary			*6/8 (75%)*	*5/7 (71%)*	*0/1 (0%)*	*3/8 (38%)*	*3/7 (43%)*	

Blank cells = not analysed or not reported. NS = not significant. X = significant. [1](uh+ .27), NS (um- .19), [2]Differences with filled pauses only occurred within phrases, [3](men+, women-), [4]Likely significant without Bonferroni correction. [5]significant differences in opposite direction younger > older. [6]Age analyzed as a linear variable. [7]there were two younger adult groups in the study with 26 participants in each.

James, 2014; Cooper, 1990; Davidson et al., 2003; Duchin & Mysak, 1987; Horton et al., 2010; James et al., 2018; Kemper et al., 2003, 2009; Martin, 2017; Schmitter-Edgecombe et al., 2000; Shewan & Henderson, 1988; Spieler & Griffin, 2006). Table 1 contains an overview of the results of these studies, and also, methodological details about sample size and the speaking task used in each study. For the total disfluencies produced (i.e., the overall rate of disfluency), 75% of the studies showed a (significant) positive result in which older adults produce more disfluency compared to younger adults. Likewise, three-quarters of all studies found significantly higher rates of filled pauses (*um's* and *uh's*) among older adults. For overall disfluency rates and filled pauses, the effect sizes were small-to-medium. There are no (significant) positive results with unfilled (silent) pauses. Finally, repetitions and repairs show significant differences in approximately one-third to one-half of the studies. Again, effect sizes for significant findings range from small-to-medium.

In our opinion, there are three key studies on disfluency and aging (Bortfeld et al., 2001; Castro & James, 2014; James et al., 2018). These studies are important because they were focused primarily on disfluency production in aging and assessed disfluency according to standard psycholinguistic convention. Bortfeld et al. (2001) investigated a large range of different factors (situational and demographic) in task-oriented conversations. They found higher rates of disfluency when speakers were discussing abstract figures, which they associated to planning difficulty. In addition, older adults produced numerically (but not significantly) more disfluency than did younger and middle-aged adults. Castro and James (2014) found that older adults produced increased disfluency when describing negative pictures, as compared to neutral pictures. In contrast, younger adults' disfluencies did not vary by picture type, and thus, those authors reported a significant interaction between age group and picture type. James et al. (2018) examined picture descriptions in which pictures varied based on whether they contained errors or not. They also reported that older adults produced more fillers for error pictures compared to non-error pictures, and that older adults were more disfluent overall (i.e., repetitions and repairs). Thus, there was a significant interaction between age group and picture type for fillers, and a main effect of age on repetitions and repairs.

To summarize, the literature on disfluency in aging contains some replication issues (particularly for repetitions and repairs). There are some possible explanations for these mixed/conflicting findings, but at present, it is not clear what those explanations are. There may be sampling issues, which result in differences in language and cognitive abilities between groups. For example, there might be differences between older adults recruited from a university setting compared to those recruited from the community. There may also be differences in the size of the speech sample analyzed for each participant. The reason that this may be an issue is that it opens the possibility of small speech samples resulting in extreme values. This would be most obvious in cases of secondary analysis (e.g., switchboard corpus), in which the researcher has little-to-no control over the data. Third, there may be some differences based on the speaking task, as there may be production demand variability. For example, comparisons of fully interactive dialogue versus situations in which the speaking situation was more "scripted" or constrained (as in many psycholinguistic sentence production and picture-naming tasks) may be confounded by task demands. For simplicity, we refer to these more scripted speaking situations, as *monologue* tasks. However, classification of speaking task into monologue vs. dialogue is a clear oversimplification of the variability of production tasks used in existing in existing literature.

Individual differences in disfluency production

The second research goal in the current study focused on examining how differences in disfluency production may be related to cognitive individual differences. Executive functions are often described as *cognitive control* mechanisms, which support a wide range of everyday behavior, particularly in the service of achieving goals (Miyake et al., 2000). They include updating/monitoring of working memory, set shifting (also called mental flexibility), and inhibitory control (Burgess, 1997; Denckla, 1996; Friedman & Miyake, 2004; Miyake & Friedman, 2012; Miyake et al., 2001). It is widely assumed that these executive functions have a role in most, if not all cognitive tasks, including language production (Roelofs, 2003). Studies examining the role of executive functioning in more complex cognitive tasks often use *individual differences* paradigms. That is, if an ability, such as language production, relies on a particular type of executive functioning (e.g., inhibition), then individuals varying in that executive function will also vary in language production.

Several studies have reported significant relationships between speech disfluencies and executive functions (e.g., Engelhardt et al., 2010, 2011, 2012, 2013). The most robust findings with respect to inhibition were shown for repair disfluencies. The inhibition-repair association was specifically shown by performance on the Stroop task and stop-signal reaction time, which accounted for nearly one-third of the variance in repair disfluencies. In contrast, repetitions were consistently linked to verbal intelligence, as measured by Wechsler Adult Intelligence subscales (i.e., vocabulary, comprehension, information, and similarities). In all cases, there were "negative" relationships between executive functioning and intelligence and disfluency production (i.e., better performing/ higher ability individuals produced fewer disfluencies).

In later work, Engelhardt et al. (2019) examined the role of working memory on speech disfluencies by examining a sentence-repetition task. The prediction driving that study was that individual differences in working memory ability, and specifically, slow retrieval or increased likelihood of retrieval failure, would be related to increased rates of disfluency (Baddeley & Logie, 1999). Confirming that prediction, results from hierarchical structural equation modeling showed that repairs were related to individual differences in working memory, and specifically, that working memory accounted for approximately one-quarter of the variance in disfluency production. Again, the direction of the relationship was such that higher ability individuals produced fewer repairs.

We identified only one study in the literature that considered executive functioning with respect to aging and disfluency production. Kemper et al. (2009) assessed disfluency production in two different speaking conditions (i.e., a baseline task and a dual task). They also assessed inhibitory control (with the Stroop task), processing speed, and working memory, and reported the correlations between filled pauses and these measures. In the baseline speaking condition, there were no significant correlations. However, the strongest correlations were with working memory, and for both younger and older adults, correlations were negative (−.17 and −.21, respectively). (Again, negative indicates that higher working memory span individuals produced fewer filled pauses.) In the dual-task condition, there was a significant correlation in younger adults between filled pauses and inhibition (i.e., .48), where higher Stroop scores indicated worse performance. (In this case, a positive relationship indicates that poorer inhibitory control related to increased rates of filled pauses.) In contrast, the correlation between filled pauses and inhibition among older adults was also positive, but not significant (i.e., .23). The likely explanation for these correlations is that participants who had difficulty dealing with dual-task demands were more likely to produce filled pauses.

There are some additional points worth mentioning about this study. The first is that filled pause rates also included interjections,[1] and therefore, represented more disfluency than just filled pauses. The second is that the study did not consider other types of disfluency (e.g., repetitions or repairs). Considering these factors, the study does not shed much light on the relationship between disfluency, executive functioning, and aging. However, the results of the study were unusual in that it is one study in the literature that showed a "reversed" effect (i.e., that younger adults produced more filled pauses than did the older adults).

Current study

The current study had two research goals. The first was to examine age differences in disfluency production and the second was to investigate the role of cognitive individual differences in disfluency production. As reviewed above, the disfluency literature with respect to aging is highly mixed. The most consistent findings for aging and disfluency concern overall disfluency rates, and therefore, we predicted that older adults would be more disfluent (total disfluencies). In classic psycholinguistic literature, filled pauses have been argued to be a more "listener-oriented" form of disfluency (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Lake et al., 2011). That is, speakers produce filled pauses as a signal to the listener that they are experiencing some difficulty and/or planning what to say, and that there will be a delay (Clark, 1994; Smith & Clark, 1993). Further, Clark and Fox Tree (2002) argued *um's* signal upcoming long delays and *uh's* signal upcoming short delays. In a more recent review paper on filled pauses, Corley and Stewart (2008) concluded (1) that filled pauses occur in situations where the speaker experiences difficulty and (2) that filled pauses provide information to listeners, who can use them to facilitate understanding. In short, they stated speakers signal "pay attention, I'm in trouble here and the next part of the message might not be what you expected" (pg. 600). It is also clear that listeners can productively use the presence of filled pauses to facilitate comprehension (e.g., Bailey & Ferreira, 2003; Fraundorf & Watson, 2011, 2013).

In the current study, we investigated more nuanced hypotheses for filled pauses, and those hypotheses were based on something called the *um-uh* ratio (for a review, see Engelhardt, 2020). The *um-uh* ratio has become increasing prominent in disfluency research, and it essentially predicts differential roles for *um* and *uh* (Horton et al., 2010; Horváth, 2010). *Um's* are thought to be more prominent discourse markers (i.e., signaling upcoming long delays – consistent with Clark & Fox Tree, 2002), whereas *uh's* are primarily produced as more local markers of difficulty experienced by the speaker. Thus, *um's* and *uh's* may have different distributions. To investigate these hypotheses, we calculated overall rates of filled pauses, the rate of *um* production, the rate of *uh* production, and the *um/uh* ratio.

There were two empirical datapoints motivating the examination of *um's* and *uh's* separately. The first is the findings of Kemper et al. (2009), who reported that younger adults produced more filled pauses than did older adults. The second is an aging study by Horton et al. (2010), who found a significant positive correlation between *uh's* and age (r =.27) (i.e., more *uh's* produced by older adults), and a significant negative correlation between *um's* and age ($r = -.19$) (i.e., fewer *um's* produced by older adults). Horton et al. also speculated that the discrepancies were due to the different functions between the two types of filled pauses. Thus, the analysis of *um's* and *uh's*, and *um/uh* ratio is more fine-grained than most prior aging-related disfluency studies, and is important theoretically for theories of speaker- vs. listener-oriented disfluency (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Lake et al., 2011).

Findings for repetitions and repairs have been more varied in the literature, with approximately one-third to one-half of studies showing a positive significant effect (see Table 1). We expected older adults to produce more repetitions and repairs. However, whether those differences are significant or not is difficult to predict given prior findings.

With respect to the second research goal, participants completed a battery of cognitive tasks, which were selected to give broad coverage to executive functioning and intelligence (Ardila et al., 2000; Carroll, 1993; Deary, 2001; Hasher & Zacks, 1988; Miyake & Friedman, 2012). We chose one measure for each of the main types of executive functioning (working memory, inhibition, and set shifting), and two measures of intelligence (vocabulary: a component of verbal intelligence, and matrix reasoning: a component of fluid intelligence). In general, prior findings of individual differences and disfluency indicated negative relationships between individual differences variables and disfluency production (i.e., higher ability individuals produced less disfluency). Based on the individual difference findings described above, we hypothesized that rates of filled pauses would be most closely associated with working memory (based on the pattern in Kemper et al., 2009 – baseline condition, and Engelhardt et al., 2019). Part of the rationale for this prediction is based on speaking task. Participants completed a story re-telling task, and thus, the task involved a clear memory component. Furthermore, we hypothesized that repetitions would be significantly related to vocabulary, which is the most consistent finding in the disfluency literature, which also assessed individual differences (see Supplementary Materials, section A for a summary of prior studies). For repairs, we hypothesized that they would be related to (1) inhibitory control (as shown in several earlier sentence-production tasks) and (2) working memory (given that the task requires memory encoding and retrieval; Markostamou & Coventry, 2021). If the inhibition predictions are confirmed, then it would provide theoretical support for the *Inhibition Deficit Hypothesis* (Hasher, 2015), as inhibitory processes are less efficient in older adults (Hasher & Zacks, 1988; Hasher et al., 2007; Lustig et al., 2007). Relatedly, if the working memory predictions are confirmed, then it would suggest that much of the variance in disfluency production can be associated with memory capacity issues, and theoretically, this would support models of language production assuming a prominent role for executive functioning (Roelofs, 2003).

Method

Participants

Participants were 154 adults with a mean age of 55.49 years (*SD* = 20.74, range 18–85), of which 63 were male and 91 were female. Participants were recruited from the University of East Anglia and the local Norwich area through opportunity sampling. All research procedures were ethically approved by the University of East Anglia's School of Psychology Ethics Committee and were carried out in accordance with the 2013 Declaration of Helsinki. Most young adults were recruited from undergraduate and postgraduate university programs through an online system and university advertisements, and were awarded course credits. All other participants were recruited from the community through advertisements in local media outlets and invitation leaflets, and received monetary compensation for their participation.

All participants spoke English as their first language, and had normal or corrected-to-normal vision and hearing. Exclusion criteria for all participants included prior history of head injury, alcohol and drug dependence, severe learning or intellectual disability, any active medical or neuropsychological condition resulting in cognitive dysfunction, and a formal subjective memory complaint (i.e., had sought professional assessment due to concerns about their memory). Inclusion criteria for participants aged 45 or older included a score ≥ 25 on the Montreal Cognitive Assessment (Nasreddine et al., 2005), a brief screening test of general cognitive functioning.

For the age group analyses, participants were divided into three groups (younger adults: 18–26 years, middle-aged adults: 45–65 years, and older adults: 66–85 years; see Figure 1 and Table 2). The groups differed largely as expected with respect to individual differences. In short, there were significant age differences in the expected direction (older < younger) for backward digit span, Stroop, trails, and matrix reasoning, and the opposite pattern (older > younger) for vocabulary (see Supplementary Materials, section B for scatterplots).

Standardizedmeasures

Inhibition

To assess inhibition, we used the Stroop task (Golden, 1978; Stroop, 1935). Participants had to name the color of the ink in which color words were printed, which was different from the written words (incongruent). Performance was based on the number of correct responses given within 45 seconds, thus, higher scores indicate better performance.

Set shifting

Set shifting (or mental flexibility) was measured using the Trail Making Test (Reitan, 1958; Strauss et al., 2006). This task consisted of two parts. In Part A, the participant had to connect in ascending order a set of numbers using a pencil, as fast as possible. In Part B, participants had to rapidly connect a series of numbers and letters in ascending order while alternating between numbers and letters (i.e., 1-A, A-2, then 2-B, etc.). In case of an

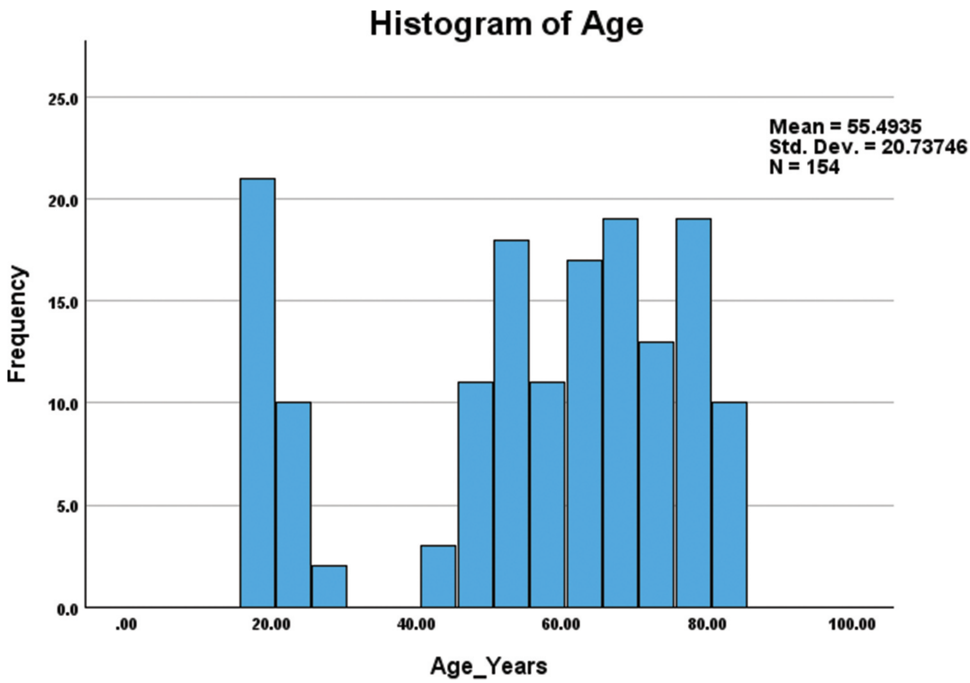

Figure 1. Histogram showing the number of participants by age. As can be seen, there were a number of younger adults (i.e., university students), and a number of middle-aged and older adults (i.e., ranging from 45 to 85 years of age).

error, participants were notified and encouraged to return to the point before the error and retrace their steps. Time to complete Part A was subtracted from time to complete Part B, hence, higher scores indicate worse performance.

Working memory

To examine working memory, participants completed the forward and backward digit span tasks (Wechsler, 2008). In these tasks, participants are provided with a random series of numbers which they have to immediately repeat either in the same order (forward digit span) or in the reversed order (backward digit span). Higher scores indicate better performance.

Fluid and verbal intelligence

The Matrix Reasoning test (Wechsler, 2008) was used to assess non-verbal fluid intelligence, as it requires making use of current information within a novel problem solving and reasoning setting (i.e., selecting the option that completes an incomplete matrix of geometric figures out of six options). Performance was based on the number of correct responses given within 5 minutes. For the assessment of verbal intelligence, we used the Mill Hill vocabulary test (Raven, 1998), which requires selecting the most relevant synonym out of six options for each word provided. It thus examines the individual's ability to utilize verbal information, providing an indication of verbal intelligence. There was no time limit for this task, and higher scores indicate better performance.

Table 2. Demographic data, intelligence, and executive functions scores broken down by age group, including statistical results.

	Young (N = 33)	Middle (N = 60)	Older (N = 61)	One-way ANOVA	Independent Samples T-tests		
Demographic	Mean(SD)	Mean(SD)	Mean(SD)		Young-Middle	Middle-Older	Young-Older
Age (years)	20.85(2.10)	55.17(5.85)	74.56(5.54)	$F(2,151) = 1162.81$***	$t(91) = -32.41$***	$t(119) = -18.74$***	$t(92) = -53.37$***
Gender (% male)	45.5	38.3	41.0	$\chi^2(2) = .45$, $p = .80$			
Education (yrs.)	13.85(1.97)	15.15(3.37)	12.69(3.08)	$F(2,151) = 10.14$***	$t(91) = -2.03$*	$t(119) = 4.19$***	$t(92) = 1.96$
Executive Functions							
Forward Digits	10.88(2.32)	11.66(2.53)	10.59(2.47)	$F(2,151) = 2.62$, $p = .08$			
Backward Digits	8.18(2.30)	7.88(2.26)	6.97(1.87)	$F(2,151) = 4.48$*			
Stroop	55.85(10.83)	43.45(9.32)	34.23(7.48)	$F(2,151) = 62.34$***	$t(91) = 5.79$***	$t(119) = 6.01$*	$t(92) = 2.77$**
Trails B-A	29.58(15.31)	34.35(16.15)	47.90(25.84)	$F(2,151) = 10.85$***		$t(119) = -3.45$**	$t(92) = -3.73$***
Intelligence							
Matrix	19.88(2.64)	18.98(3.34)	15.80(4.23)	$F(2,151) = 18.09$***		$t(119) = 4.59$***	$t(92) = 5.03$***
Vocabulary	17.94(3.49)	22.61(4.13)	23.44(3.76)	$F(2,151) = 23.34$***	$t(91) = -5.51$***		$t(92) = -6.94$***

*$p < .05$, **$p < .01$, ***$p < .001$. Data based on un-transformed means.

Speech samples

To collect speech samples, participants were told four novel stories (see Supplementary Materials, Section C), which they had to memorize and repeat back to the experimenter as accurately as possible (Markostamou & Coventry, 2021; Redford, 2013). Participants had to repeat these stories twice, once immediately after the stories had been read to them, and once again after an interval of approximately 25 minutes. For the second repetition, participants were given a short prompt before each story. The stories produced by participants were recorded, and afterward transcribed and coded for disfluency.

General procedure

Upon entering the lab, participants provided written informed consent and basic demographic information. They then completed each of the tests in the battery (verbal and non-verbal intelligence tests, digit span tasks, trail making test, Stroop task, and the story re-telling task). The same experimenter (second author) conducted individual testing sessions with each participant. The entire testing session lasted approximately 2 hours.

Disfluency coding

Three main types of disfluency were examined: filled pauses, repetitions, and repairs.[2] Filled pauses are *uh's* or *um's*, and there were a few instances of *er* and *erm*. Repetitions refer to repetitions of a word or string of words with no functional benefit. Repairs occur when a speaker stops speaking, and then starts over with a new word or phrase.

The dataset was transcribed and coded by two trained research assistants, with each completing approximately half of the data. A third trained research assistant then coded approximately half of the data (as a validation sample). Next, the results from the first team and the validation sample were compared. The results showed high correlations, suggesting good reliability (filled pauses: $r(776) = .948$, $p < .01$; repetitions: $r(776) = .847$, $p < .01$; repairs: $r(776) = .901$, $p < .01$). The first author resolved the few discrepancies that did occur, and ensured there were no systematic differences in the coding. The first author was responsible for classifying filled pauses into *um's* and *uh's*, and did so by re-listening to the entire dataset. The dataset contained 58,764 words (approximately 382 words per participant). There were no significant differences in the number of words produced based on age group or disfluency rates between the first and second re-telling of the stories (all $p's > .05$). The dependent variable was the rate of disfluency production (calculated for each type separately) per word produced by each participant. The *um/uh* ratio was calculated by dividing rate of *um* production by the sum of *um's* plus *uh's* (i.e., *um/(um + uh)*). This *um/uh* calculation is technically a proportion rather than a ratio, but we use the "ratio" label for consistency with prior literature.

Data screening and preparation

Data points greater than ± 3.0 standard deviations from the mean for each variable in the data set were scrutinized as outliers. There were none in the individual differences measures, and three in the disfluency means. The disfluency coding for these cases was re-checked, and two of the three occurred within one participant. Given the low number of elevated means, we retained them in the dataset. Finally, we applied a transformation to skewed variables (i.e., square root and inverse).

Data analytic plan

The data analysis plan consisted of two main sections. In the first section, we compared rates of disfluency production among the three groups of younger, middle-aged, and older adults, using a one-way ANOVA. This was done for the total disfluencies produced and each disfluency type examined. These analyses are most comparable to existing disfluency studies of aging in the literature and relate to whether there are overall differences in disfluency production based on age. This section addresses the first goal of the study.

The second main section focused on the individual differences variables. We began this section by presenting the correlations between variables. Second, we present the results of six backwards regressions (one for each type of disfluency – total filled pauses, *um's, uh's, um/uh* ratio, repetitions, and repairs). For these regressions, we also included age as a continuous variable.[3] These analyses examine which individual differences variables significantly relate to rates of disfluency production. We chose regression for these analyses, because they reveal whether each variable accounts for "unique" variance in the dependent variable (disfluency), and thus, regression provides a more thorough assessment compared to examining simple bi-variate correlations.

Results

The descriptive statistics and bivariate correlations between variables are presented in Tables 3 and 4, respectively.

Table 3. Descriptive statistics for intelligence, executive function measures, and disfluencies.

Measure	*N*	Mean	SD	Min	Max	Skew	Kurtosis
Executive Functions							
Forward Digit Span	154	11.05	2.50	5.0	16.0	0.20	–.53
Backward Digit Span	154	7.58	2.17	3.0	14.0	0.80	0.76
Stroop	154	42.45	12.01	15.0	76.0	0.48	0.29
Trails B-A	154	38.69	21.69	6.0	121.0	1.36	2.32
Intelligence							
Matrix	154	17.92	3.98	7.0	26.0	−0.78	0.26
Vocabulary	154	21.94	4.38	11.0	32.0	−0.25	−0.40
Proportion of Disfluencies							
Total Disfluent	154	0.53	0.32	0.0	0.17	1.01	1.35
Filled Pauses	154	0.03	0.03	0.0	0.14	1.33	2.73
Um rate	154	0.02	0.02	0.0	0.13	1.80	5.99
Uh rate	154	0.01	0.01	0.0	0.11	3.15	14.43
Um/uh ratio	154	0.62	0.31	0.0	1.00	−0.60	−0.63
Repetitions	154	0.01	0.01	0.0	0.04	1.61	3.02
Repairs	154	0.01	0.01	0.0	0.03	0.70	0.70

Data based on un-transformed means.

Table 4. Bivariate correlations between age, executive function and intelligence measures, and disfluencies.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age	-	−.08	−.22**	−.71**	.36**	−.42**	.48**	−.13	−.22**	−.32**	.04	−.21**	.20*	−.06
2. Forward Digit		-	.57**	.29**	−.27**	.18*	.24**	.06	.07	.06	.05	.07	.01	.02
3. Backward Digit[a]			-	.46**	−.46**	.36**	.18*	.01	.09	.07	.09	.06	−.12	.10
4. Stroop				-	−.49**	.49**	−.18*	.02	.08	.22**	−.13	.27**	−.27**	.07
5. Trails B-A[a]					-	−.54**	−.05	.07	.00	−.06	.10	−.14[#]	.25**	−.09
6. Matrix Reasoning						-	.09	−.05	−.03	−.04	.02	.01	−.20*	−.05
7. Vocabulary							-	−.19*	−.20*	−.25**	.03	−.07	−.03	.07
8. Total Disfluent[a]								-	.91**	.77**	.61**	.29**	.53**	−.50**
9. Filled Pauses[a]									-	.85**	.66**	.33**	.30**	−.21*
10. *Um* rate[a]										-	.22**	.67**	.24**	−.15[#]
11. *Uh* rate[a]											-	−.28**	.21**	−.14[#]
12. *Um/uh* ratio[a]												-	.09	−.06
13. Repetitions[a]													-	−.44**
14. Repairs[b]														-

[#]$p < .10$, $^{*}p < .05$, $^{**}p < .01$, [a]square root transformation, [b]inverse transformation. Highlighted correlations show the significant correlations between individual differences variables and disfluency measure. Shaded cells indicate the significant relationships between individual differences variables and disfluency.

Section 1 – disfluency rates by age group

We began the analysis by investigating differences in disfluency production based on *age group*. There were no significant differences in the mean number of words produced (younger = 386, middle-aged = 379, older = 382) by participants in the different groups $F(2,150) = .07$, $p = .93$, $\eta^2 = .00$. In addition, the total disfluencies produced were not significantly different between groups $F(2,151) = 1.19$, $p = .31$, $\eta^2 = .02$. The trend was for the younger participants to produce more disfluencies overall (i.e., the mean rate of disfluency production for younger participants was .059), which was driven primarily by the rates of filled pauses. In contrast, the middle-aged and older adults produced less disfluency overall (i.e., means of .053 vs. .050, respectively).

Disfluency types

Results showed that there were significant group differences in the number of total filled pauses $F(2,151) = 3.68$, $p = .03$, $\eta^2 = .05$. Younger adults produced more filled pauses than did middle-aged adults $t(91) = 2.13$, $p = .02$. There were also significant differences between younger and older adults $t(92) = 2.67$, $p = .01$. However, the difference between middle-aged and older adults was not significant $t(119) = .63$, $p = .53$ (see Figure 2, left panel).

The means for filled pauses broken down by type are also shown in Figure 2 (right panel). Results showed significant group differences in the number of *um's* $F(2,151) = 7.45$, $p < .001$, $\eta^2 = .09$. Younger adults produced more *um's* than did middle-aged adults $t(91) = 2.71$, $p = .01$. There were also significant differences between younger and older adults $t(92) = 3.81$, $p < .001$. However, the difference between middle-aged and older adults was not significant $t(119) = 1.28$, $p = .20$. In contrast, there were no differences in the rate of *uh* production $F(2,151) = .093$, $p = .91$, $\eta^2 = .001$. For *um/uh* ratio, results showed that the difference between groups was marginally significant $F(2,151) = 2.60$, $p = .08$, $\eta^2 = .03$ (younger adults = .73 (SD = .25), middle-aged adults = .59 (SD = .32), and older adults = .58 (SD = .32)). Following up the marginal result, there were significant differences between the younger and middle-aged adults $t(91) = 2.09$, $p = .04$ and between the younger and older adults $t(92) = 2.21$, $p = .03$. The middle-aged and older adults were not significantly different $t(119) = .23$, $p = .82$. What the analysis of filled pauses shows is

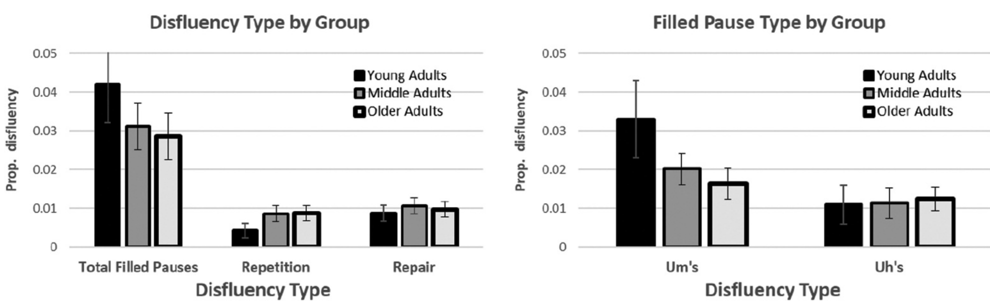

Figure 2. Proportion of disfluencies (per word) produced. Left panel shows results for total filled pauses, repetitions, and repairs. Right panel shows results for *um's* and *uh's*. Error bars show the 95% confidence intervals.

that younger adults produce more *ums* than do the middle-aged and older adults, and this translates to a marginally significant difference in the *um/uh* ratio.

There were also significant age group differences on the number of repetitions $F(2,151) = 4.87$, $p = .01$, $\eta^2 = .06$. However, for this analysis, both the middle-aged and older adults produced more disfluencies than did the younger adults [middle-aged vs. younger: $t(91) = -3.10$, $p = .003$, and older vs. younger: $t(92) = -2.69$, $p = .01$]. In contrast, there was no difference between the middle-aged and older adults $t(119) = .20$, $p = .84$. Finally, for repairs, the groups were not significantly different from one another $F(2,151) = 1.08$, $p = .34$, $\eta^2 = .014$.

To summarize, results showed that younger adults produced more filled pauses than did the middle-aged and older adults, and specifically, they produced more *um's*. For repetitions, the pattern reversed. Younger adults produced fewer repetitions than did middle-aged and older adults. Repairs and *uh's* showed no significant group differences.

Section 2 – individual differences variables

The correlations presented in Table 4 showed that the individual difference variables produced several significant correlations with disfluency rates. Total filled pauses were significantly related to age and vocabulary. *Um's* were significantly related to age, Stroop, and vocabulary. *Uh's* were not significantly related to any variables. *Um/uh* ratio was significantly related to age and Stroop. Repetitions were significantly related to age, Stroop, trails, and matrix reasoning. Repairs were not significantly related to any variables. In order to investigate how these variables were related to the production of disfluency, we ran six backwards regressions (one for each type of disfluency and *um/uh* ratio) with the purpose of identifying which variables significantly related to disfluency production. We also included age in the regression analyses.

Results of the regression analyses are presented in Table 5. Assumptions of regressions were checked (i.e., residuals were normally distributed and equal distribution of error

Table 5. Regression coefficients for retained predictor variables by disfluency type.

Variable	*B*	*SE (B)*	*β*	*t*-value
Total Filled Pauses $F(1,152) = 7.54$, $p = .01$, $R^2 = .047$				
Age	−.001	.000	−.22	−2.75**
Um's $F(2,151) = 12.40$, $p < .001$, $R^2 = .141$				
Age	−.001	.000	−.41	−4.96**
Matrix	−.004	.001	−.21	−2.53*
Uh's $F(2,151) = 3.33$, $p = .04$, $R^2 = .042$				
Backward	.029	.014	.18	2.04*
Stroop	−.001	.000	−.21	−2.34*
Um/Uh ratio $F(2,151) = 7.58$, $p < .001$, $R^2 = .091$				
Stroop	.008	.002	.35	3.89**
Matrix	−.012	.006	−.16	−1.83#
Repetitions $F(2,151) = 7.59$, $p < .001$, $R^2 = .091$				
Stroop	−.001	.000	−.19	−2.16*
Trails B-A	.005	.003	.16	−1.77#
Repairs $F(3,150) = 1.87$, $p = .14$, $R^2 = .036$				
Age	.000	.000	−.22	−2.07*
Matrix Reasoning	.000	.000	−.16	−1.71#
Vocabulary	.000	.000	.19	1.96#

***$p < .001$, **$p < .01$, *$p < .05$, #$p < .10$.

variance), and we also assessed multi-variate outliers via Cook's distance. All Cook's D values were acceptable (i.e., < .12).

Age was significantly related to total filled pauses, which is consistent with results in Section 1. Younger participants produced more total filled pauses than did the older participants. Likewise, age was a predictor of *um's*, and was in the same direction as it was for total filled pauses. However, matrix reasoning was also a predictor of *um's*. The direction of the effect was such that individuals with higher matrix reasoning showed a lower rate of *um* production. For *uh* production, there were two significant variables retained in the regression. They were backwards digit span and the Stroop task. For backwards digit span, higher memory ability participants produced a greater number of *uh's*. For the Stroop task, individuals with greater inhibitory abilities produced fewer *uh's*. The *um/uh* ratio showed that the Stroop task was significant and matrix reasoning was marginally significant. Individuals with better Stroop performance showed a higher ratio (i.e., higher ratio of um-to-uh), and for matrix reasoning, the direction showed that higher ability individuals had a lower ratio (i.e., lower ratio of uh-to-um).

For repetitions, two variables were retained (i.e., the Stroop task and the trails task). Stroop performance was related to repetitions, such that individuals with higher inhibitory ability showed significantly fewer repetitions. The trails task was related to repetitions, such that individuals with higher set shifting ability showed marginally fewer repetitions. Three additional points are worth mentioning here. First, age was not a significant predictor of repetitions (in contrast to results in Section 1). Second, matrix reasoning, which showed a significant bi-variate correlation with repetitions, was also not significant. We believe that this is likely due to the "shared" variance between matrix reasoning and the executive function measures. Both correlations were very near |.50|.

For repairs, three variables were retained (age, matrix reasoning, and vocabulary). However, it is important to note that the overall regression model for repairs was not significant, consistent with the fact that the correlations showed no significant relationships. In addition, two of the three retained predictors were only marginally significant.

Discussion

This study had two research goals. The first focused on age-related differences in disfluency production. As reviewed in the Introduction, the literature on disfluency production in aging is extremely mixed, and we speculated on several reasons why this may be the case. The results of this study showed significant age differences in filled pauses, specifically in the rates of *um* production. We also observed significant differences in the rates of repetitions. The second research goal focused on the role of cognitive individual differences in rates of disfluency. For this goal, we found that several significant results for executive functions and intelligence.

Age differences in disfluency production

The first goal examined disfluency production by comparing disfluency rates across three age groups (younger, middle-aged, and older adults). We predicted that older adults would produce significantly more disfluency. Prior literature was especially conflicting for repetitions and repairs with approximately one-third to one-half of prior studies showing a significant result. However, based on two studies (Horton et al., 2010; Kemper et al., 2009), we further examined **type** of filled pause by classifying them into *um's* and *uh's*, and we also calculated the *um/uh* ratio. Our predictions were that *um's* and *uh's* would have slightly different distributions, and if there were differences, it would shed light on the listener- vs. speaker-oriented forms of disfluency.

The results of this study, with respect to age, showed that younger adults produced significantly more total filled pauses than middle-aged and older adults, counter to predictions (see also, Kemper et al., 2009). This was due to their increased tendency to produce *um's* (with an effect size of .09), whereas the production of *uh's* did not show a significant age-group difference (see also Horton et al., 2010). The increased *um* production in younger adults carried through in the *um/uh* ratio differences between groups, but it is important to note that *um/uh* ratio was only marginally significant. The fact that younger adults produced more *um's* suggests an increased tendency to signal their delays than do middle-aged and older adults. To place the *um* and *uh* rates in this study in context, the *um/uh* ratio for younger adults in the current study (.73) was almost identical to controls in ASD literature (.77), whereas middle-age and older adults in the current study have a lower ratio (.58–.59). Individuals with ASD have an even lower *um/uh* ratio (.47) (Engelhardt, 2020).

The age analysis further showed that there were significant differences in repetitions, with middle-aged and older adults producing more repetitions compared to younger adults (with and effect size of .06), consistent with predictions. More specifically, younger adults produced about half the number of repetitions as did the middle-aged and older adults, which is consistent with approximately one-third of the existing studies in the literature. The rate of repairs was not significantly different in the age group comparisons, and again, the null effect of repairs is consistent with approximately one-third of the studies in the literature.

To summarize the findings of this study with respect to age differences, we observed that younger adults produced more *um's* than did middle-age and older adults, and that younger adults produced fewer repetitions than did middle-age and older adults.

Cognitive individual differences and disfluency production

The second goal of this study focused on the relationship between disfluency production and cognitive individual differences. We ran a series of backwards regressions, in which we assessed whether any of the individual difference variables were significantly related to rates of disfluency. The general prediction for this set of analyses was that higher performing (or higher ability) individuals would produce fewer disfluencies. The specific predictions were that filled pauses would be related to working memory, repetitions would be related to vocabulary, and repairs would be related to both working memory and inhibition. If confirmed, these findings would lend theoretical support for both theories of aging (e.g., Inhibition Deficit Hypothesis) and models of language production, which assume a role for executive functioning (e.g., WEAVER ++).

Filled pauses

For total filled pauses, age was the only significant predictor, which is consistent with the age-group results above. Age was also retained as a significant predictor in the rate of *um* production. Younger adults produced more *um's* than did the middle-age and older adults, suggesting that younger adults are better at signaling delays in their speech. This difference in *um's* drove the significant difference for total filled pauses. Matrix reasoning was also retained in the regression examining *um* production. The direction of the effect was such that higher ability individuals produced fewer *um's*. Recall that based on Horton et al. (2010), we hypothesized that *um's* would have a distinct distribution compared to *uh's*, and be produced as a signal (or discourse marker, as argued by Horton et al.) of an upcoming delay. The matrix reasoning result clearly suggests that *um's* (or at least a good proportion of *um's*) reflect problems experienced in the course of speaking rather than a "helpful" form of listener-oriented disfluency. One additional point to note about *um* production is that vocabulary and Stroop resulted in significant correlations but neither was retained in the final regression model, and we believe that the lack of an effect for Stroop is due to the fact that age was a stronger predictor, and age and Stroop had a lot of overlapping variance (i.e., the highest correlation in the entire dataset).

The retained predictors for *uh's* were backward digit span and Stroop. The direction of the Stroop result is consistent with predictions (i.e., better Stroop performance corresponded to fewer *uh's*). In contrast, backward digit span was not consistent with general predictions, as higher digit span corresponded with higher rates of *uh* production. A possible explanation for this result is that individuals with better memory abilities are actually able to recall more details of the stories (a sort of tip of the memory phenomenon), and thus, they produce *uh's* in the process of retrieving and conveying these details. In order to investigate this explanation, we ran a correlation between backward digit span and the number of words produced. The correlation was not significant ($r = .13$, $p = .10$), but positive, which indicates a numerical trend for higher ability individuals to recall more information about the story. By this explanation, higher ability individuals actually expend greater effort, which in turn makes it more likely that they produce this one type of disfluency (i.e., *uh's*). Finally, the regression for *um/uh* ratio showed that Stroop and matrix reasoning were retained, and the direction of Stroop and matrix reasoning were the same as the two previous analyses, and thus, consistent with predictions.

Repetitions

We predicted that repetitions would be related to verbal intelligence (e.g., Engelhardt et al., 2010). In the current study, however, we found that Stroop performance was significantly related to repetitions, and the trails task was marginally related to repetitions. Thus, reduced inhibitory control seems to result in production problems, such that, something "upcoming" in the speech stream is problematic, forcing speakers to stop speaking. When they resume, they start over with something already articulated (i.e., prior to the point of suspension). This is thought to be an attempt to restore continuity to an interrupted constituent (Clark & Wasow, 1998). There is one other study in the literature that also reported a significant correlation between inhibition (Stroop) and repetitions (Engelhardt et al., 2013). The correlation in that study was −.19, which is virtually identical to the standardized regression coefficient in the current study (see Table 5). These findings are consistent with the Inhibition Deficit Hypothesis. The marginal effect of the trials task makes is difficult to make firm conclusions about the role of set shifting in repetitions, and we note also that the bi-variate correlation was also marginally significant.

Repairs

The overall regression model for repairs was not significant ($p = .14$). Moreover, the bi-variate correlations with repairs showed no significant effects (see Table 4). Thus, we do not believe that there is much discussion given the null findings.

Summary

Virtually all of the significant individual differences effects were consistent with our general predictions that higher ability individuals would produce less disfluency. We did not observe a working memory effect on either filled pauses or repairs, nor did we observe a vocabulary effect on repetitions. In fact, the only significant memory effect found was in the opposite direction of predictions. Instead, matrix reasoning was significantly related to *um's* and marginally related to repairs. We found that inhibition (Stroop performance) was significantly related to two types of disfluency (*uh's* and repetitions). Both of which are consistent with one of the main theories of aging (i.e., the Inhibition Deficit Hypothesis). More generally, these cognitive individual difference findings will also be important to theories of language production which assume a role for executive functioning.

Strengths, limitations, and future directions

The main strengths of the present study are (1) the large sample size and (2) the extensive battery of tasks completed by all participants, including measures of short-term/working memory, inhibition, set shifting, vocabulary, and matrix reasoning. This kind of individual differences study and the size of the speech sample generated from each participant represents a substantial empirical undertaking. The study does have several limitations. The first and most obvious is that the sample had an age-hole in it. We did not recruit individuals in their 30s, and also, the number of individuals in the late teens and 20s was smaller than those in the middle-age and older adult groups. This potentially means that individual younger participants had a greater impact on statistical analyses. However, because the range of ages was 18 and 85, and there was a large number overall, we do not feel that the study is sufficiently impacted by the missing 30 year olds, nor the comparatively smaller "younger" age group. Cognitive declines are also not typically observed in this age range. (The correlations for participants 45 years and older are presented in the Supplementary Materials, section E.) The second limitation is the cross-sectional nature of the study, which may be an issue if there are cohort differences in patterns of speech. This limitation can only be addressed by future longitudinal studies, which would more definitively demonstrate causal relationships between individual differences and the tendency to produce disfluent speech. The third limitation is that education was not controlled between groups (see Table 2). In particular, the middle-aged group had a higher number of years of education compared to the younger and older groups. We did run the correlations between education (years) and the disfluency measures, and none were significant. However, it would be desirable to have had our groups matched in terms of education.

We think the one obvious future direction concerns rates of disfluency production in atypically aging individuals, including individuals with mild cognitive impairment or mild dementia due to Alzheimer's disease (Kemper et al., 2001). One possibility is that the early linguistic declines associated with dementia may also encompass an increase in disfluency production (Kemper & Lyons, 1994). If so, disfluencies may provide an additional early clinical-diagnostic marker of cognitive impairment/decline. A second broad future direction is longitudinal studies, which would be better placed to establish a causal connection between declines in cognitive individual differences and increased disfluency with age (see Beirer et al., 2023).

Conclusions

The present study does not answer all of the questions about the mixed findings in the literature or definitively show **why** older adults produce higher rates of certain disfluency types. Instead, what the current study contributes several nice pieces to the puzzle regarding (1) disfluency production in aging and (2) the role of individual differences in disfluency production (in aging). That is, taken as a whole the present study contributes broadly to the disfluency literature. At the same time, the current study addresses one theoretical reason **why** older adults, in some cases, produce more disfluency. The Inhibition Deficit Hypothesis assumes a prominent role of decrements inhibitory control on language function across the lifespan, and some of our results are clearly in line with this theoretical prediction (Hasher, 2015). Both *uh's* and repetitions were related to Stroop task performance (Friedman et al., 2006).

To summarize the main findings of the present study, we found that younger adults produced more filled pauses, and specifically, more *ums* than did the older participants. Higher rates of *um* production were likely due to better signaling of an upcoming delay (see also Horton et al., 2010). Differences in *uh's* were linked to lower inhibition and better memory performance. Repetitions were linked to inhibitory control and set shifting. When individual differences were tested in statistical models, age was no longer significantly related to repetitions. We believe that the current results will be important to future theoretical work in which the role of individual differences are further explicated and modeled (Bock & Levelt, 1994; Dell & O'Seaghdha, 1992; Dell et al., 1999; Kempen & Huijbers, 1983; Levelt, 1989; Vosse & Kempen, 2000). The findings lead to a range of exciting new research questions, for example, does inhibitory control affect particular stages of production, such as, lemma selection or phonological encoding?

Notes

1. Interjections are inserted linguistic material (e.g., *I mean, You know, well, etc*.), which is unrelated to the main message being conveyed. These are also sometimes referred to as *lexical* fillers.
2. We also coded blends. However, these were infrequently produced, and therefore, excluded from all analyses.

3. We also checked whether disfluency rates varied by gender (Martin, 2017). Those results are presented in the Supplementary Materials, section D. We also performed a series of correlational analyses between education (in years) and the disfluency measures, and none were significant.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ioanna Markostamou http://orcid.org/0000-0001-7343-0122

Data availability statement

The authors will make the dataset available following publication.

References

Albert, M. L. (1980). Language in normal and dementing elderly. In L. K. Obler & M. L. Albert (Eds.), *Language and communication in the elderly: Clinical, therapeutic, and experimental issues* (pp. 145–150). D.C. Health.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology*, *15*(1), 31–36. https://doi.org/10.1093/arclin/15.1.31

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200. https://doi.org/10.1016/S0749-596X(03)00027-5

Beirer, E. J., Chantavarin, S., & Ferreira, F. (2023). Do disfluencies increase with age? Evidence from a sequential corpus study of disfluencies. *Psychology and Aging*, *38*(3), 203–218. https://doi.org/10.1037/pag0000741

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in spontaneous speech: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Psychology Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579. https://doi.org/10.1016/0749-596X(91)90026-G

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Sciences*, *13*(1), 21–24. https://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* (pp. 373–443). Psychology Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Castro, N., & James, L. E. (2014). Differences between young and older adults' spoken language production in descriptions of negative versus neutral pictures. *Aging, Neuropsychology & Cognition*, *21*(2), 222–238. https://doi.org/10.1080/13825585.2013.804902

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication*, *15*(3–4), 243–250. https://doi.org/10.1016/0167-6393(94)90075-2

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242. https://doi.org/10.1006/cogp.1998.0693

Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology*, *45*(5), P210–P214. https://doi.org/10.1093/geronj/45.5.P210

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, *2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Davidson, D. J., Zacks, R. T., & Ferreira, F. (2003). Age perseveration of the syntactic processor in production. *Journal of Psycholinguistic Research*, *32*(5), 541–566. https://doi.org/10.1023/A:1025402517111

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517–542. https://doi.org/10.1207/s15516709cog2304_6

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314. https://doi.org/10.1016/0010-0277(92)90046-K

Denckla, M. B. (1996). A theory and model of executive function: A neuropsychological perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263–278). Brookes.

Dennis, P. A. (2012). *Aging and word retrieval: How aging-related gains in semantic knowledge and loss of access to linguistic information are expressed in connected speech*. North Carolina State University.

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*(3), 245–257. https://doi.org/10.1016/0021-9924(87)90022-0

Engelhardt, P. E. (2020). Speaker- versus listener-oriented disfluency. In F. R. Volkmer (Ed.), *Encyclopedia of autism spectrum disorders* (pp. 4531–4540). Springer.

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition*, *38*(5), 617–628. https://doi.org/10.3758/MC.38.5.617

Engelhardt, P. E., Ferreira, F., & Nigg, J. T. (2011). Language production strategies and disfluencies in multi-clause network descriptions: A study of adult attention-deficit hyperactivity disorder. *Neuropsychology*, *25*(4), 442–453. https://doi.org/10.1037/a0022436

Engelhardt, P. E., McMullon, M. E. G., & Corley, M. (2019). Individual differences in the production of disfluency: A latent variable analysis of memory ability and verbal intelligence. *Quarterly Journal of Experimental Psychology*, *72*(5), 1084–1101. https://doi.org/10.1177/1747021818778752

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica*, *144*(2), 424–432. https://doi.org/10.1016/j.actpsy.2013.08.002

Engelhardt, P. E., Veld, S., Nigg, J. T., & Ferreira, F. (2012). Are language production problems apparent in adults who no longer meet diagnostic criteria for attention-deficit/Hyperactivity disorder? *Cognitive Neuropsychology*, *29*(3), 275–299. https://doi.org/10.1080/02643294.2012.712957

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167. https://doi.org/10.1016/S0010-0277(96)00781-0

Fraundorf, S. H., & Watson, D. G. (2011). The disfluency discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, *65*(2), 161–175. https://doi.org/10.1016/j.jml.2011.03.004

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*(9), 1083–1096. https://doi.org/10.1080/01690965.2013.832785

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Golden, C. J. (1978). *Stroop color and word test: A manual for clinical and experimental issues*. Stoelting.

Gordon, J. K., Anderson, K., Perez, G., & Finnegan, E. (2019). How old do you think I am? Speech-language predictors of perceived age and communicative competence. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2455–2472. https://doi.org/10.1044/2019_JSLHR-L-19-0025

Gordon, J. K., & Kindred, N. K. (2011). Word retrieval in ageing: An exploration of the task constraint hypothesis. *Aphasiology*, *25*(6–7), 774–788. https://doi.org/10.1080/02687038.2010.539699

Goulet, P., Ska, B., & Kahn, H. J. (1994). Is there a decline in picture naming with advancing age? *Journal of Speech and Hearing Research*, *37*(3), 629–644. https://doi.org/10.1044/jshr.3703.629

Hasher, L. (2015). Inhibitory deficit hypothesis. In S. K. Whitbourne (Ed.), *The encyclopaedia of adulthood and aging* (pp. 1–5). John Wiley & Sons, Inc.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). Oxford University Press.

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). Academic Press.

Heine, M. K., Ober, B. A., & Shenaut, G. K. (1999). Naturally occurring and experimentally induced tip-of-the-tongue experiences in three adult age groups. *Psychology and Aging*, *14*(3), 445–457. https://doi.org/10.1037/0882-7974.14.3.445

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*(3), 708–713. https://doi.org/10.1037/a0019424

Horváth, J. (2010). "Discourse features", syntactic displacement and the status of contrast. *Lingua*, *120*(6), 1346–1369. https://doi.org/10.1016/j.lingua.2008.07.011

James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1378–1391. https://doi.org/10.1037//0278-7393.26.6.1378

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*(3), 355–367. https://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 younger and older adults. *Aging, Neuropsychology & Cognition*, *25*(4), 520–534. https://doi.org/10.1080/13825585.2017.1337061

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Kemper, S. (1992). Adults' sentence fragments who, what, when, where, and why. *Communication Research*, *19*(4), 444–458. https://doi.org/10.1177/009365092019004003

Kemper, S., Herman, R., & Lian, C. (2003). Age differences in sentence production. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(5), P260–P268. https://doi.org/10.1093/geronb/58.5.P260

Kemper, S., Kynette, D., Rash, S., O'Brein, R., & Sprott, K. (1989). Life-span changes to adults' language: Effects of memory and genre. *Applied Psycholinguistics*, *10*, 49–66. https://doi.org/10.1017/S0142716400008419 (1)

Kemper, S., & Lyons, K. (1994). The effects of Alzheimer's disease on language and communication. In M. L. Hummert, J. Wieman, & J. Nussbaum (Eds.), *Interpersonal communication and older adulthood: Interdisciplinary theory and research* (pp. 58–82). London: Sage.

Kemper, S., & Rash, S. (1988). Speech and writing across the life-span. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 107–112). John Wiley & Sons.

Kemper, S., Schmalzried, R., Herman, R., Keedahl, S., & Mohankumar, D. (2009). The effects of aging and dual task demands on language production. *Aging, Neuropsychology & Cognition*, *16*(3), 241–259. https://doi.org/10.1080/13825580802438868

Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, *16*(4), 600–614. https://doi.org/10.1037/0882-7974.16.4.600

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging*, *21*(4), 804–809. https://doi.org/10.1037/0882-7974.21.4.804

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review*, *18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

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

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view. In D. In: Gorfein & C. MacLeod (Eds.), *The place of inhibition in cognition* (pp. 145–162). American Psychological Association.

Markostamou, I., & Coventry, K. (2021). Memory for route and survey descriptions across the adult lifespan: The role of verbal and visuospatial working memory resources. *Journal of Environmental Psychology*, *78*, 101712. https://doi.org/10.1016/j.jenvp.2021.101712

Markostamou, I., & Coventry, K. R. (2022). Naming spatial relations across the adult lifespan: At the crossroads of language and perception. *Neuropsychology*, *36*(3), 216–230. https://doi.org/10.1037/neu0000789

Martin, F. M. (2017). Vocabulary, grammar, sex, and aging. *Cognitive Science*, *41*(4), 950–975. https://doi.org/10.1111/cogs.12367

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640. https://doi.org/10.1037/0096-3445.130.4.621

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*(8), 1217–1238. https://doi.org/10.1080/17470210701467912

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Raven, J. (1998). *Ravens progressive matrices and vocabulary scales*. Psychological Corporation.

Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics*, *34*(3), 569–589. https://doi.org/10.1017/S0142716411000877

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and word-finding: A comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology*, *15*(6), 479–493. https://doi.org/10.1093/arclin/15.6.479

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: The network dynamics of cognitive decline and perseveration. *Science*, 583–587.

Shewan, C. M., & Henderson, V. L. (1988). Analysis of spontaneous language in older normal population. *Journal of Communication Disorders*, *21*(2), 139–154. https://doi.org/10.1016/0021-9924(88)90002-0

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38. https://doi.org/10.1006/jmla.1993.1002

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*(1–3), 291–321. https://doi.org/10.1080/01690960400002133

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration norms and commentary*. Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Verhaegen, C., & Poncelet, M. (2013). Changes in naming and semantic abilities with aging from 50 to 90 years. *Journal of the International Neuropsychological Society*, *19*(2), 119–126. https://doi.org/10.1017/S1355617712001178

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143. https://doi.org/10.1016/S0010-0277(00)00063-9

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). Pearson.

Wierenga, C. E., Benjamin, M., Gopinath, K., Perlstein, W. M., Leonard, C. M., Rothi, L. J. G., Conway, T., Cato, M. A., Briggs, R., & Crosson, B. (2008). Age-related changes in word retrieval: Role of bilateral frontal and subcortical networks. *Neurobiology of Aging*, *29*(3), 436–451. https://doi.org/10.1016/j.neurobiolaging.2006.10.024