Article:

Hanif, Naveen, Jefferies, Elizabeth orcid.org/0000-0002-3826-4330 and de Bruin, Angela (2024) Express:Naming speed during language production in younger and older adults: Examining the effects of sentence context. Quarterly journal of experimental psychology (2006). 17470218241309602. ISSN 1747-0226

https://doi.org/10.1177/17470218241309602

Naming speed during language production in younger and older adults: Examining the effects of sentence context

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD-24-263.R1
Manuscript Type:	Standard Article
Date Submitted by the Author:	22-Oct-2024

Complete List of Authors:	Hanif, Naveen; University of York Jefferies, Elizabeth; University of York, Psychology de Bruin, Angela; University of York

Keywords:	Language production, cognitive ageing, semantic control, sentence context

DOI: 10.1177/17470218241309602

Naming speed during language production in younger and older adults: Examining the effects of sentence context

Naveen Hanif[1], Elizabeth Jefferies[1], & Angela de Bruin[1]

[1]University of York

Corresponding author:
Angela de Bruin
Department of Psychology
University of York
York
YO10 5DD
angela.debruin@york.ac.uk

DOI: 10.1177/17470218241309602

Abstract

Word retrieval during speech production has been found to slow down with ageing. Usually, words are produced in sentence contexts. The current studies examined how different sentence contexts influence lexical retrieval in younger and older adults. We also examined the potential influence of semantic knowledge and control on sentence-context effects. Study 1 was completed by 48 younger and 48 older adults. They named pictures that were preceded by a matched context (which predicted that specific target word), a mismatched context (predicting another word), a neutral context (that did not predict one specific word), or no context. In comparison to the neutral context, both younger and older adults' word production was faster in matched contexts, suggesting both age groups benefited from sentence contexts facilitating the retrieval of predictable words. Neither age group was slowed down by the mismatched contexts (compared to the neutral contexts), suggesting these contexts did not create (sufficient) interference to hinder lexical retrieval. In Study 2, participants completed measures of semantic knowledge, verbal fluency, semantic control, and inhibition. Older adults showed larger semantic knowledge but poorer inhibition and (on some measures) semantic control than younger adults. However, none of these measures predicted the sentence context effects observed in Study 1. Together, this suggests older adults' lexical retrieval can continue to benefit from sentence contexts predictive of upcoming words during language production.

Keywords: Language production, cognitive ageing, semantic control, sentence context

DOI: 10.1177/17470218241309602

1. Introduction

The number of people aged 65 and over is increasing rapidly (ONS, 2021). Ageing has consequences for language production, including lexical retrieval. This is the process through which words stored in the mental lexicon are accessed for production (cf. Friedmann et al., 2013). Lexical retrieval difficulties are the most common type of memory difficulty experienced by older adults (e.g., Ossher et al., 2013). In particular, the efficiency (speed) with which words are retrieved from memory declines with age (Verhaegen & Poncelet, 2013). In the existing literature, lexical retrieval has most commonly been studied in isolation (e.g., during picture naming). However, in everyday conversation, language is often used to respond to other people, for example when they ask you a question. The sentence context (e.g., the nature of the question asked by our conversational partner) might influence how quickly we can retrieve the words we need to respond to them. Across two studies, we therefore examined lexical retrieval efficiency in younger and older adults, in response to different types of sentence contexts. We also examined how cognitive factors (particularly those known to be affected by ageing) influenced the younger and older adults' ability to respond to the different sentence contexts.

1.1. Lexical retrieval in older adults

Picture-naming tasks often show that older adults (typically defined as aged 65 years or older) have more difficulty naming objects compared to younger adults (e.g., Barresi et al., 2000; Connor et al., 2004; Goral et al., 2007; MacKay et al., 2005). While accuracy can sometimes be preserved in older adults (Boudiaf et al., 2018; Ferré et al., 2020; Hoyau et al., 2017; Schmitter-Edgecombe et al., 2000), or only show age-group differences after the age of 70 (Wen & Dong, 2023), naming times appear more sensitive to age-related changes at a relatively earlier age (e.g., Ferré et al., 2020; Hoyau et al., 2017). For instance, slower naming compared to younger adults has been observed from the age of fifty (e.g., Verhaegen & Poncelet, 2013). Indeed, while the vocabulary itself (i.e., the knowledge) remains intact or even increases with age (e.g., Hoffman, 2018), it is the (speed of) *retrieval* of words that appears impacted. This can affect low-frequency words in particular (e.g., Ferré et al., 2020).

These changes in lexical retrieval can be explained through several cognitive changes generally observed with ageing, including general age-related slowing (e.g., Salthouse, 1996) and difficulties suppressing goal-irrelevant information (e.g., Hasher & Zacks, 1988). Focusing on language specifically, the transmission deficit hypothesis explains these findings through weakened connections between representations at different levels within the ageing lexicon (Burke et al., 1991). While semantic information about words can remain intact, this account argues that weaker connections from the lexical to the phonological level can result in poorer lexical retrieval in older adults.

DOI: 10.1177/17470218241309602

1.2. Language in context

Daily-life lexical retrieval typically takes place in context, including sentences and interaction with other people. Studies examining connected speech in older and younger adults have shown mixed findings regarding age-related differences. Some studies report that older adults perform more poorly in connected speech tasks than younger adults do (e.g., older adults produce fewer words and make more word-choice errors, Heller & Dobbs, 1993). Other studies have reported similar performance in older and younger adults on several measures (e.g., in terms of the number of words produced and disfluencies during neutral picture descriptions; Castro & James, 2014) or show that older adults produce more words than younger adults during connected speech (e.g., Arbuckle et al., 2000; see Kavé & Goral, 2017 for a review). Based on their review, Kavé and Goral (2017, p.521) concluded that *"there is little evidence for significant word retrieval deficits in connected speech production in healthy aging."* Furthermore, performance on picture-naming and connected speech tasks does not always correlate (e.g., Saling et al., 2012; Schmitter-Edgecombe et al., 2000). This might in part be related to the different measures used. Connected speech studies have used a range of variables, including number of words produced, lexical diversity (variety in the words used), word substitutions and circumlocutions (suggesting failed target-word retrieval), coherence, and dysfluencies (e.g., Kavé & Goral, 2017; Hoffman et al., 2018). Picture-naming tasks, however, often measure *speed* of lexical retrieval (naming times) and/or accuracy. These measures might tap into different aspects of language production and thus make it difficult to directly compare retrieval in context (connected speech) versus in isolation (picture naming). Naming times measured during picture naming might be most sensitive to detecting earlier and smaller changes in lexical retrieval (e.g., Verhaegen & Poncelet, 2013). Furthermore, some measures used in connected speech might partly assess *compensatory strategies* (Kavé & Goral, 2017), for example in the form of circumlocutions that can mask retrieval difficulties (e.g., Nicholas et al., 1985). The various measures used across studies hinder a direct comparison, making it difficult to evaluate whether age-related effects on lexical retrieval and language production are truly influenced by the presence of context. In Study 1, we therefore assessed speed of picture naming (naming times) when pictures were presented in isolation and when they were preceded by a sentence context.

As well as the *presence* of context, the *type* of context might influence lexical retrieval. Words are often retrieved faster when preceded by contexts with lexical-semantic information that is compatible with the target word. For example, following the question "*What is he woken up by every morning?"* the picture of an *alarm clock* is named faster than after a semantically neutral question such as *"What did he hear yesterday?"* (Shao & Rommers, 2020). Upon hearing or seeing a word, activation of its semantic and/or lexical features can spread activation to neighbouring

DOI: 10.1177/17470218241309602

representations that share features or associations. This priming can facilitate production of related words, compared to unrelated words that do not share semantic features. This has also been linked to prediction, with listeners argued to predict upcoming words, potentially through pre-activation of specific lexical features that are likely to appear (e.g., Federmeier et al., 2007). In contexts where those predictions are accurate, this could facilitate the speaker's own production. However, in daily-life speech and language, we often also encounter words that are not highly predictable (Luke & Christianson, 2016). Contexts that are not compatible with a target word (i.e., where the prediction is incorrect) can result in processing costs (e.g., Federmeier et al., 2003, but cf. Luke & Christianson, 2016). This could potentially slow down language production, although a recent study (Bannon et al., in press) suggested this might especially be the case when the unexpected words are not related to the predicted target word at all.

These semantic relationships can further modulate age-related differences in lexical retrieval as older adults have shown poorer semantic control, the mechanism through which intended representations (semantic knowledge) are retrieved from the semantic store while competing representations are supressed (Hasher & Zacks, 1988; Jefferies, 2013). While older adults' semantic knowledge is often reported to be comparable to, or even larger than, younger adults' knowledge (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003), semantic control and inhibitory control diminish with age (e.g., Hasher & Zacks, 1988; Hoffman, 2018; Spieler et al., 1996). Weakened inhibitory control can create difficulties in suppressing responses and information irrelevant to the task at hand. The term "inhibitory control" is often used as a general term to refer to control over different types of information, both linguistic/verbal and non-linguistic/non-verbal. Semantic control specifically refers to control over irrelevant semantic representations (e.g., distractor words that are semantically related to the target while having to make a size judgement). Both semantic and inhibitory control have been found to predict some aspects of word production (Arbuckle and Gold, 1993; Hoffman et al., 2018; Yin & Peng, 2016; but see Higby et al., 2019). Age-related changes in semantic networks and control processes may thus have implications for lexical retrieval processes in everyday conversation.

Our Study 1 therefore first examined lexical retrieval in older and younger adults in different sentence contexts that varied in their semantic relationships with the target word to be named. The preceding sentence was either neutral (not semantically related to the target), predicted the target word ("matched", e.g., "mountain" was predicted by "What did the alpinist climb?"), or predicted a different target ("mismatched", e.g., "ladder" rather than the target "mountain" was predicted by "What did the construction worker climb?"). In Study 2, we examined whether lexical retrieval in these different contexts related to performance on other cognitive tasks, including semantic control and

DOI: 10.1177/17470218241309602

inhibition. Together, these studies allowed us to examine potential age-related changes in lexical retrieval during word production in sentence context, as well as the underlying mechanisms contributing to lexical retrieval in context. Specifically, considering the observed age-group differences in previous studies, our studies also aimed to better understand the potential impact of (age-related changes in) semantic knowledge and control on language production.

2. Study 1

2.1. Introduction

2.1.1. Sentence context

Previous literature has mostly assessed the role of sentence context in older adults through comprehension studies, with little work assessing production. Production has mostly been studied through priming paradigms manipulating the relationship between an individual prime word and target word (e.g., "doctor-nurse"). Both younger and older adults show faster naming when target words are preceded by a semantic prime compared to a neutral prime (Balota et al., 1999; Faust et al., 2004), with some findings suggesting older adults may benefit even more from semantic priming than younger adults (see Laver & Burke, 1993 for a review). The *comprehension* studies that have looked at word processing in sentence contexts have reached a range of conclusions. Some have found that both younger and older adults use sentence context to facilitate retrieval of upcoming words to the same extent (e.g., Kliegl et al., 2004, using eye tracking). Other findings suggest that older adults can utilise semantic cues to a *greater* degree than younger adults (e.g., Pichora-Fuller et al., 1995; Rayner et al., 2006), lending support to the hypothesis that older adults can utilise semantic context to help them to overcome age-related declines (see also Rayner et al., 2006; Speranza et al., 2000). However, other studies suggest that older adults do not use semantic information within sentential contexts as effectively as younger adults do. These findings are often shown through EEG data examining N400 effects, a negative-going wave peaking approximately 400ms after stimulus onset. This N400 effect is often reduced in amplitude in congruent sentence contexts, but such sentence context effect has been observed to be smaller or delayed for older than younger adults (e.g., Federmeier et al., 2002; Wlotko et al., 2012).

Together, these findings suggest older adults can continue to use semantically congruent information during language processing, although it remains unknown whether they can benefit more or less than younger adults. These findings are based on studies looking at language comprehension (e.g., sentence processing), leaving it also largely unknown how congruent ('matched") sentence contexts can influence language production in older adults.

DOI: 10.1177/17470218241309602

Furthermore, the focus has been on facilitation stemming from semantically congruent information, but in contexts that are incongruent with a target word, older adults might experience more interference based on diminished inhibitory and/or semantic control. Studies looking at semantic interference at the individual word-level (e.g., naming the picture of a *ball* while seeing the word *frisbee*) have shown both younger and older adults show slower naming in the presence of this distractor than when they see a neutral word. This semantic interference effect is often greater within older adults than in younger adults (e.g., Taylor & Burke, 2002), suggesting that older adults have more difficulty in inhibiting semantic distractors. However, this is not found across all studies, with Lorenz et al. (2019) showing a similar impact of semantic distractors on younger and older adults. Some ERP studies, finally, have suggested that older adults might be influenced less by incongruent sentence contexts as a consequence of predicting upcoming words less or less successfully (e.g., Wlotko et al., 2012). However, these studies often compare unexpected, incongruent words to expected, congruent words without a neutral baseline. This makes it difficult to disentangle effects of (potentially facilitating) congruent contexts and (potentially interfering) incongruent contexts.

2.1.2. Rationale Study 1

The current literature thus has shown mixed effects regarding age-related differences in terms of lexical retrieval in context. It has furthermore focused on comprehension and, in the absence of a neutral baseline, often does not allow for a comparison between semantically congruent (matched) and incongruent (mismatched) contexts. Study 1 therefore examined the effects of both congruent (matched) and incongruent (mismatched) sentence contexts (relative to a neutral baseline) on lexical retrieval in younger and older adults. We used a picture-naming paradigm similar to Shao and Rommers (2020), who showed faster naming in younger adults when naming a picture after a matching question that was related to that target word, compared to a neutral question. In addition, we also included a mismatching sentence context as well as a picture-naming task without context (see Table 1 for example stimuli).

In terms of our hypotheses, we first expected a “Match effect”, with faster picture naming after a matched than neutral sentence. Based on previous literature, it was unclear if older adults’ Match effect would be similar to that of younger adults. Preserved or increased semantic knowledge in older adults (cf. Hoffman, 2018) might help both age groups equally to retrieve words in matched sentence contexts, or might help older adults even more than younger adults to compensate for slower lexical retrieval (larger Match effect). However, less efficient use of semantic information (for example due to slower transmission between representations or less prediction forming, e.g., Dell, 1986; Federmeier et al., 2010) may result in a smaller or no Match effect for older adults.

DOI: 10.1177/17470218241309602

Furthermore, we expected a "Mismatch effect" with slower naming in semantically mismatched contexts than in neutral contexts. In line with the inhibition deficit hypothesis (Hasher & Zacks, 1988) and decreased semantic control in older age (Hoffman, 2018), semantically mismatched information may be more likely to interrupt older adults' retrieval of upcoming words (larger Mismatch effect). A similar Mismatch effect in younger and older adults would suggest that mechanisms used to inhibit competing words are not negatively affected by ageing. Finally, if older adults do not use semantic information to predict upcoming words to the same extent as younger adults, they might experience less interference (smaller Mismatch effect).

Finally, we aimed to examine whether context in general (producing picture names in response to a neutral question) can influence lexical retrieval, relative to no context (producing individual picture names). This was done in an attempt to overcome the issues faced previously when trying to compare word production across different tasks employing differing measures (Kavé & Goral, 2017). Age effects might be exacerbated in a relatively artificial task asking participants to produce individual words without the typical syntactic and lexical connections between words in context. Faster retrieval *within* a sentence context (compared to an isolated word) would suggest that those syntactic and lexical connections between words (as is common in daily-life speech) can aid lexical retrieval, even if the context does not provide clear semantic predictions. If older adults rely more heavily on contextual support to aid word retrieval, they may exhibit a larger context effect. On the other hand, listening and responding to another speaker may impose greater working memory demands than producing words without context. If older adults have more difficulty managing the working memory demands of producing words within conversation, they may be delayed by context (Kemtes & Kemper, 1997; Murphy et al., 2000).

--- Insert Table 1 about here ---

2.2. Methods

This study was pre-registered on the Open Science Framework: https://osf.io/8qexr/. The data for both studies are also available on that OSF page.

2.2.1. Participants

Ethical approval was obtained from the Department of Psychology at the University of York and participants provided informed consent at the start of the study. The final sample included 96 native English-speaking monolinguals. Forty-eight older adults (aged 65-77 years old) were recruited through

DOI: 10.1177/17470218241309602

prolific.co (n=34) and through our departmental database (n=14). Forty-eight younger adults (aged 18-35) were recruited through SONA (the university's internal participant recruitment system; n=6) and Prolific (n=42). Participants received either monetary compensation, Amazon vouchers, or course credit for their participation. The groups of younger and older adults were matched on sex ratio, number of years of formal education received, and the number of participants within each age group who had completed at least an undergraduate degree (see Table 2).

Participants were first asked to complete a series of checks, including testing their microphone and playing audio files. We checked these responses before inviting them to the full study. Three participants were not invited as they could not complete the pre-study checks; one participant did not respond to the invitation; and three participants completed the study but were not included as they did not follow the instructions or because their naming-task recordings were empty. Our final sample size of 96 participants (as pre-registered, and after exclusion) was based on a GPower analysis. It was not possible to retrieve previous effect sizes from the literature as comparable designs/tasks had not been used previously with younger and older adults. We therefore conducted a power analysis using a medium effect size (f = 0.25) for an interaction between age and sentence context. This suggested our sample size yielded over 95% power to detect a medium-sized effect.

All included participants furthermore confirmed meeting the following eligibility requirements: they did not use a hearing aid, had (corrected-to-)normal vision, were not colour blind, had not been using medication affecting their concentration in the past three months, did not have a language/reading disability, and had not been diagnosed with a neurodegenerative disease or cognitive impairment. Given that the study was conducted online, we were not able to use an assessment of cognitive functioning such as the ACE-III. In addition to asking participants to confirm each eligibility point, where possible, we also used existing screening criteria to only invite participants without a history of head injury, cognitive impairment, or dementia.

--- Insert Table 2 about here ---

2.2.2. Design

Participants completed a picture-naming task with the within-participant independent variable Context. This had four levels (see Table 1 for examples): Matched, Mismatched, Neutral, or No Context. Age group was a between-subject variable. The dependent variable was picture-naming times (ms), defined as onset of naming relative to picture presentation.

DOI: 10.1177/17470218241309602

2.2.3. Materials

All target pictures for the naming task were presented in greyscale and were sourced from the Multipic database (Dunabeitia et al., 2018) or from Google images. Pictures were preceded by a spoken question or presented without context. The questions were recorded by a female English speaker, reflecting natural speech as much as possible. They were pre-processed using Praat (Boersma & Weenink, 2022) to add 50ms to the beginning and end of each recording and to scale all recordings to 60dB. Background noise was also reduced using Audacity® version 3.0.0.

We created 76 matched question-answer pairs, in which the question was strongly predictive of the upcoming picture (see Table 1). Each matched pair was combined with another matched pair to create a duo (e.g., in Table 1, "egg" and "rattle" form a duo). Duos were formed on the basis that the sentence formed a match with one target word but a mismatch with the other target in the duo. We created mismatch sentences in which the target word was unlikely to follow but not impossible, to avoid unrealistic scenarios that would never happen in real-life conversations. Each duo was also assigned a neutral question, which did not strongly prime a specific word.

Each participant named each picture four times: three times within context (once per matched, mismatched, and neutral question) and once without context. We ensured that participants only heard each question once so that they could not use previous exposure to predict a word. Using the example presented in Table 1, half of the participants named "egg" four times in the four conditions while the other half named "rattle" four times in the same contexts. Matched and mismatched contexts were therefore the same questions across participants. Neutral questions were matched to the matched/mismatched questions in terms of overall sentence length (number of words) and syllable length and frequency of the key words. The full list of stimuli, with further details about the sentence characteristics and matching, is provided in the online Supplementary Materials.

To make sure the stimuli functioned as intended (i.e., target words were most likely in matched contexts and least likely in mismatched contexts), we ran three pilot studies, as described below. The pilot studies were completed online with our initial set of stimuli. Participants were recruited through SONA, Qualtrics, and Prolific for all three pilots.

The pilot studies included 47 sets of initially prepared stimuli (comprising n = 47 each of matched, mismatched and neutral questions). Changes to the stimuli were made based on the pilot responses. The final set of stimuli was also evaluated through a likeliness rating task in the main study. This confirmed that target words were most likely in matched contexts and least likely in mismatched contexts (see "Results" for an analysis of these ratings).

One pilot study was a short written-picture naming task and was completed by five older (*M* Age = 63.6 years, range = 61-68) and six younger adults (*M* Age = 18.7 years, range = 18-20) to make

DOI: 10.1177/17470218241309602

sure all pictures could be recognised and named easily. We replaced pictures where this was not the case. Two further pilot studies were conducted to examine suitability of the stimulus materials in the different contexts. In the first, 21 older (*M* Age = 65.7 years, range = 60-75 years) and 20 younger adults (*M* Age = 19.85 years, range = 18-31) completed a cloze probability task and a likeliness rating task. In the cloze probability task, they viewed each question individually and were asked to generate their first three single-word answers in response. We computed cloze probabilities (i.e., the proportion of times the target word was given as part of that "top three") through by-item means rather than by-participant means, and below report scores only including items that were kept in the same form in the actual experiment. Cloze probability was highest for the matched condition in both age groups (Younger: *M* = 80.52%, *SD* = 24.22; Older *M* = 84.56%, *SD* = 19.77). This confirmed the matched target responses were indeed good answers to the questions. In contrast, as we wanted, the target words were almost never given in the mismatched context (Younger *M* = 2.38%, *SD* = 6.80; Older *M* = 2.78%, *SD* = 9.15) and neutral context (Younger *M* = 1.72%, *SD* = 5.51; Older *M* = 1.48%, *SD* = 5.65).

--- *Insert Table 3 about here* ---

Participants in the pilot also completed a likeliness ratings task. They viewed each question-answer pair and were asked to rate on a scale of 1-5 how *likely* the presented answer was to follow the preceding question. Ratings (only including items that were kept in similar form in the actual experiment) were highest for matched pairs, followed by neutral pairs, and were lowest for mismatched pairs (see Table 3). A 2x3 ANOVA confirmed that there was a significant effect of Condition ($F(1.663,84.831) = 199.591$, $p < .001$), with a significant difference between Matched and Mismatched ($p < .001$), Matched and Neutral ($p < .001$), and between Mismatched and Neutral ($p = .015$) likeliness ratings in the expected direction. Older adults overall provided slightly higher ratings ($F(1,51) = 38.832$, $p < .001$), which interacted with Condition ($F(1.718,87.615) = 18.964$, $p < .001$). This reflected that while younger and older adults rated likeliness of matched pairs similarly, older adults' ratings of neutral and mismatched pairs were slightly higher than the younger adults' ratings. However, analyses by age group confirmed that likeliness ratings were highest for the matched and lowest for the mismatched question-answer pairs in each age group individually.

Finally, we asked participants to indicate whether the scenario depicted in each question-answer pair was possible (i.e., whether it was something that could happen in real life, as we wanted the mismatched sentences to be unlikely but not impossible) and whether the question-answer pairs were grammatically correct. For both questions, the answer options were "yes" (can happen in real

DOI: 10.1177/17470218241309602

life/grammatical) or "no" (cannot happen in real life/not grammatical). Both plausibility and grammaticality judgements (inferred from the mean number of participants who agreed that the scenarios could happen in real life and that they were grammatically correct) were high for all question-answer pairs we included.

After making modifications to the stimuli based on the results from the first pilot studies, we then conducted another pilot study with another five younger (*M* Age = 22.8 years, range = 19-29) and older adults (*M* Age = 63 years, range = 60-67). Scores for the modified stimuli again, in line with pilot described above, confirmed that the targets were most likely to follow matched questions and least likely to follow mismatched questions.

2.2.4. Procedure

The experiment was conducted using Gorilla.sc (Anwyl-Irvine et al., 2020). Participants first read the information sheet and provided informed consent. They then completed a background questionnaire (see "Participants"). Next, they completed a sound check to ensure that they could record audio files through their browser and to adjust their device's volume so that they were able to clearly hear the sentences. For the naming task, participants were allocated to one of twelve experiment lists. Half of the participants named the Set 1 targets and the other half the Set 2 targets (see the online Supplementary Materials). Furthermore, half of the participants named the pictures with context first, while the other half named the pictures without context first. Participants named 38 target words four times each, once without context and once in each of the three question contexts (114 trials, with a break in the middle). The presentation order of the stimuli was pseudo-randomised in the context blocks so that the same word was not repeated twice in a row and there were no more than three consecutive trials of the same type of context.

Participants first completed a picture familiarisation task in which they saw the target pictures and words, asking them to read the word aloud and use it during the task. This phase was included to make sure all participants recognised the pictures when naming them in the study. Given that pictures were repeated within the task across the four conditions, including a familiarisation phase ensured participants did not see the picture for the first time within the main task, which could have affected condition comparisons. In the naming tasks, participants were instructed to name the pictures as quickly and accurately as possible. Participants first saw three practice trials. In the Context blocks, participants first viewed a fixation cross (500ms) followed by a blank screen while they heard a pre-recorded question. This was followed by another fixation cross (presented for 500ms). Next, the article "the" was presented on the screen (500ms), and then another fixation cross (300ms) was presented before the picture was presented. The picture remained on screen for 2500ms, regardless of when a

DOI: 10.1177/17470218241309602

response was given. In the No-Context block, participants viewed a single fixation cross of the mean duration of the sentence recordings in the Context blocks, plus the duration of the fixation crosses (total of 3853.8 ms). The rest of the trial was identical to the Context trials (article followed by picture).

We also assessed participants' subjective experienced workload using the NASA-TLX (Task Load Index, Hart & Staveland, 1988). This task was used to examine (potential) differences in younger and older adults' experienced subjective demands (rated on a scale of 1 (very low) to 100 (very high)) during the naming task. This allowed us to assess potential age-group differences not just in terms of objective performance (i.e., RTs) but also in terms of experienced workload, which might be higher for older than younger adults. After each naming block, participants provided ratings evaluating how mentally demanding, physically demanding, and temporally demanding (pace of the block) they found the task, as well as their performance (how successful they felt in terms of following the task instructions), effort (how hard they had to work), and their frustration level. We also assessed 'overall workload' by asking participants to complete the full NASA-TLX after finishing the full naming task. This again asked participants to complete the same ratings (listed above), but we now also asked participants which aspect (e.g., "effort" versus "mental demand", asking this question for each combination of the six experiences) they found more important when describing the experienced workload. This allowed us to compute scores reflecting the participants' experienced workload per part of the task, as well as an overall score that also took into consideration that different aspects of workload experiences vary in how important they are for individual participants.

Finally, participants completed a likeliness-rating task in which they rated on a scale of 1-7 how likely the target word was to follow each question, for all of the question-answer pairs they viewed during the naming study. The experiment lasted approximately 30 to 45 minutes in total.

2.2.5. Data analysis

Likeliness ratings

Likeliness ratings in the main study were examined using a 2x3 ANOVA with Age (younger, older) as a between-subject variable and Context (matched, mismatched, neutral) as a within-subject variable. Data from two older participants were excluded from the likeliness-ratings analyses. Due to a technical fault one participant was unable to use the ratings scale to indicate a likeliness rating of greater than '3'. The other excluded participant provided a rating of '2' on all trials. Given that Mauchly's test of sphericity (for this task and all others) indicated the assumption of sphericity was violated for the Context variable, Greenhouse-Geisser corrected values are reported.

DOI: 10.1177/17470218241309602

Picture naming

An accurate response in the picture naming task was either the intended target word or a closely related word (e.g., "painter" instead of "artist"). Other or no responses were scored as an inaccurate response. Picture naming accuracy was >75% for all participants (*M* older adults 96.04%, *SD* = 4.92; *M* younger adults 97.00%, *SD* = 3.37). As pre-registered, because accuracy was close to ceiling, it was not analysed further. Naming RTs were determined using Checkvocal (Protopapas, 2007). RTs <300ms or more than 2.5 *SD* above or below the mean per participant and per condition were removed, using the trimr package (Grange, 2015; removing 2.78% of correct responses). With the exception of the data for the older adults in the Mismatch condition, Shapiro-Wilk tests confirmed the data were normally distributed (*p*s >.25).

RTs were analysed in SPSS using a 2x4 ANOVA to determine whether there was a main effect of Age, Context (i.e., matched, mismatched, neutral, without context), or an interaction between the two. If an effect of context was found, a pairwise comparison (Bonferroni corrected) was used to establish where the effect resided within the four levels. Given that we were specifically interested in the effect of each specific type of context and because we wanted to examine them while also accounting for age-related slowing, we then also computed the Match, Mismatch and Context effects based on z-scored RTs for each participant (z-scored separately per age group). The Match effect was the RT difference between the matched and neutral questions; the Mismatch effect was the difference between mismatched and neutral questions; and the Context effect was the difference between neutral questions and naming without context. In addition to a one-way ANOVA per effect (using a Bonferroni adjusted significance threshold of $p = 0.016$ to account for the three comparisons), we computed a Bayesian ANOVA using JASP version 0.17.3 (JASP Team, 2022), which examined evidence for/against an age-group difference on these contextual effects. For each contrast effect (Matched, Mismatched, Context), we compared a model including an age effect (between-groups difference) to a null model (no age-group difference). We report these results in the form of "BF_{01}", showing the evidence for the null hypotheses (no age group difference) over the alternative hypotheses (significant age group differences). Values below 1 indicate evidence for an age-group difference; values above 1 indicate evidence for no age-group difference.

Finally, we estimated the internal consistency of the Match, Mismatch, and Context effects using a permutation-based split-half approach (Parsons, 2020a) with 5000 random splits to check for within-subject variations in these effects.

Further exploratory analyses used linear mixed-effect analyses to examine the context effects while considering both participants and stimuli within one analysis. We also examined the potential role of a specific target word's likeliness scores as provided by the participants after

DOI: 10.1177/17470218241309602

completing the naming task. Finally, we examined the potential influence of lexical frequency, considering that older adults have shown increased difficulty retrieving low-frequency words. These analyses were conducted using R (4.4.1; lme4 package version 1.1.35) and started with the maximal random-effect structure including all within-participant and within-item slopes (following Barr et al., 2013). Where analyses did not converge, we removed slopes explaining the lowest amount of variance until convergence was reached. Two-level categorical variables were contrast-coded (Age group: younger adults = -0.5; older adults = 0.5). Simple coding was used for the four-level categorical variable Context. “Neutral” was used as the reference level to compare the other three levels to that baseline (contrary to dummy coding, the intercept corresponds to the mean of all cell means). The continuous variables' item frequency and item likeliness rating were z-scored.

Experienced workload (NASA-TLX)

Overall workload effects from the NASA-TLX were calculated by counting how often participants chose each experience as most important between two comparison options (e.g., how often they said they found "frustration" the most important compared to another experience in post-test evaluation). The raw score for each experience was computed and multiplied by the number of times it was chosen as most important. All weighted experiences were summed up and divided by the total number of comparisons participants had to choose from to compute the overall NASA score. We also calculated participants’ mean workload score after each block (by calculating the average of their ratings on each Likert scale, without weighing categories). An ANOVA was conducted to determine whether there was a difference in *overall* workload experience between older and younger adults. A mixed ANOVA established whether there was a significant effect of Context (ratings provided after No Context, the first half of Context, the second half of Context, and for the overall task) and age group on experienced workload.

2.3. Results

2.3.1. Likeliness ratings

--- Insert Table 4 about here ---

Target words in the matched Context were rated as most likely and targets in the mismatched Context as least likely (see Table 4, $F(1.729, 159.107) = 2327.697$, $p < .001$, $\eta_p^2 = .962$). Pairwise comparisons showed significant differences between all Context combinations ($p < .001$). There was no main effect of Age ($F(1,92) = 2.905$, $p = .092$, $\eta_p^2 = .031$), suggesting that overall ratings were similar for older and younger adults. However, there was a significant interaction between Age and Context (F(1.729,

DOI: 10.1177/17470218241309602

159.107) = 4.155, p = .022, η_p^2 = .043). While ratings of the matched and mismatched sentences were similar for both age groups (Matched: p = .718; Mismatched: p = .225), neutral targets were rated slightly more likely by younger than older adults (p = .025, see Table 4). Crucially, however, each age group showed a significant difference in likeliness ratings between all three question contexts (Younger adults: F(1.656, 77.820) = 1029.568, p < .001, η_p^2 =.956; Older adults: F(1.743, 78.441) = 1329.523, p < .001, η_p^2 =.967, with all pairwise comparisons p < .001 in both age groups). Thus, for both age groups, as intended, matched targets were most likely, followed by neutral and mismatched targets.

2.3.2. Picture naming

--- Insert Figure 1 about here ---

We started the picture-naming analysis with the untransformed RTs. As expected, there was a significant main effect of Age group on RTs, with older adults taking significantly longer to name pictures compared to younger adults (Younger M = 848.70ms, SD = 100.76; Older M = 960.19, SD = 135.03; F(1, 94) = 21.358, p < .001, η_p^2 =.185). There was also a main effect of Context on picture naming times (F(1.525, 143.337) = 110.600, p < .001, η_p^2 =.541). RTs were fastest in the matched trials, followed by the mismatched and neutral trials, and were slowest in the no-context trials (see Table 5, Figure 1). Pairwise comparisons showed that there were significant differences between RTs in all naming contexts (ps <.001 matched versus mismatched, neutral, and no-context; p = 0.005 for mismatched versus no-context; p = 0.008 for neutral versus no-context), apart from trials in the neutral and mismatched contexts (p > .999). This showed that matched contexts facilitated RTs compared to neutral contexts (Match effect) and that context facilitated RTs in comparison to naming without context. However, no Mismatch effect was observed, suggesting the mismatched context did not negatively affect production. Importantly, there was no significant interaction between Age group and Context (F(1.525, 143.337) = 1.266, p = .279, η_p^2 =.013), suggesting that the context effects did not differ between the younger and older adults.

--- Insert Table 5 about here ---

Contrast Analyses for Match, Mismatch, and Context effects

As the previous analysis showed that there was a significant effect of age group on RTs, we z-scored the data to account for age-related slowing. Then, for each participant, we computed their Match

DOI: 10.1177/17470218241309602

effect (match versus neutral RTs), Mismatch effect (mismatch versus neutral), and Context effect (neutral context versus no-context; see Figure 2).

--- Insert Figure 2 about here ---

Starting with the Match effect, both the one-way ANOVA ($F(1,94) = 0.233$, $p = .630$, $\eta_p^2 = 0.002$) and the Bayesian analysis ($BF_{01} = 4.20$, error = 0.02%) suggested the Match effect did not differ between age groups. The (Spearman-Brown corrected) split-half internal consistency of the Match effect was $r_{SB} = 0.60$, 95%CI [0.46, 0.72], indicating moderate internal consistency of this effect within participants.

Similarly, both analyses suggested the Mismatch effect did not differ between age groups either (one-way ANOVA: $F(1,94)= 0.116$, $p = .734$, $\eta_p^2 = 0.001$; $BF_{01} = 4.43$, error = 0.02%). The (Spearman-Brown corrected) split-half internal consistency of the mismatch effect was $r_{SB} = 0.19$, 95%CI [-0.10,0.43], suggesting low internal consistency of this effect within participants.

Finally, the one-way ANOVA again showed no age-group difference in the Context effect ($F(1,94)= 1.954$, $p = .165$, $\eta_p^2 = 0.020$), although the Bayesian analysis only provided weak support for this null hypothesis ($BF_{01} = 1.97$, error = 0.02%). The (Spearman-Brown corrected) split-half internal consistency of the context effect was $r_{SB} = 0.90$, 95%CI [0.86,0.93], suggesting high internal consistency of this effect within participants. While there is good internal reliability *within participants,* Figure 2 also shows high variability in the context effect across participants. Some participants exhibited context-related facilitation whilst others exhibited a context-related cost. We therefore conducted further exploratory analyses to examine if order (naming in context or without context first) could explain this variability. To that end, we computed a 2x2 ANOVA, with the Context effect as the dependent variable and Age and Naming order (completing context or no-context part first) as independent variables. There was a significant main effect of naming order ($F(1, 92) = 176.538$, $p < .001$, $\eta_p^2 = .657$). Both younger and older adults showed a facilitatory effect of context only when the No-Context part was completed first ("context second group": *M* younger = -0.745, *SD* = 0.382; *M* older = -0.603, *SD* = 0.295) but not when the Context part was completed first ("context first group": *M* younger = 0.172, *SD* = 0.348; *M* older = 0.365, *SD* = 0.360). The effect of age was now significant ($F(1, 92) = 5.584$, $p = .020$, $\eta_p^2 =.057$), suggesting the context effect was slightly smaller for older adults, but this did not interact with naming order ($F(1, 92) = 0.131$, $p = .718$, $\eta_p^2 = .001$).

DOI: 10.1177/17470218241309602

Exploratory analyses (pre-registered mixed-effect analyses exploring stimulus characteristics)

The analyses conducted so far used by-participant means. Given that language production is also influenced by the items used, we also ran a linear mixed-effect analysis. We report the analysis using untransformed RTs but also ran them using z-scored RTs, which showed the same findings (apart from the main effect of age group, as intended). The first mixed-effect analysis examined effects of Context and Age group, similar to the ANOVAs reported above, and converged with participant and item intercepts but no slopes. The Context findings were the same, with a Match effect (β = -134.128, SE = 4.103, t = -32.691, p < .001), no Mismatch effect (β = -2.836, SE = 4.127, t = -0.687, p = .492), and a Context effect (β = 42.732, SE = 4.118, t = 10.376, p < .001). The Match and Mismatch effects did not differ between age groups (Match x Age: β = -6.510, SE = 8.205, t = -0.793, p = .428; Mismatch x Age: β = 1.321, SE = 8.253, t = 0.160, p = 0.873). The Context effect was significantly smaller for older adults (β = -30.864, SE = 8.236, t = -3.747, p < .001).

Contrary to our predictions, there was no significant Mismatch effect. Likeliness ratings in the Mismatch condition were significantly lower than likeliness ratings in the other conditions, but within each condition our stimuli varied in their likeliness. We therefore also conducted an exploratory analysis to examine whether there was a direct relationship between a participant's likeliness rating for a given item and their RTs in response to that item. Across all conditions, higher likeliness ratings were associated with faster responses (β = -54.485, SE = 1.692, t = -32.198, p < .001). This was also the case when just considering the Mismatch trials (β = -12.514, SE = 3.755, t = -3.333, p < .001). None of the analyses, however, showed an interaction between likeliness and age group (all ps > .15).

A final exploratory analysis examined a potential role of word frequency, considering older adults have been found to have greater difficulty retrieving low-frequency words. No effects of frequency were observed on overall RTs (β = -5.249, SE = 6.743, t = -0.778, p = .439) and, importantly, frequency did not interact with age group (β = 0.294, SE = 2.940, t = 0.100, p = .920). Frequency did not interact with Context either (all ps > .25).

2.3.3. Experienced workload (NASA-TLX)

Subjective workload experiences showed no main effect of Age (F(1, 94) = 0.052, p = . 820, η_p^2 = .001), suggesting that younger and older adults experienced workload similarly (see Table 6). Furthermore, there was no significant effect of Context (F(2.521, 236.968) = 2.378, p = .081, η_p^2 = .025), suggesting that workload did not differ as a result of naming with or without context or throughout the context blocks. This suggests that workload was consistent throughout the naming task. There was also no significant interaction between Context and Age (F(2.521, 236.968) = 0.224, p = .848, η_p^2 = .002). A one-way ANOVA also showed that overall workload (weighted for the importance of each experience

DOI: 10.1177/17470218241309602

per participant) did not differ between younger and older adults either ($F(1, 94) = 0.055$, $p = .815$, $\eta_p^2 = .001$).

--- Insert Table 6 about here ---

2.4. Discussion

Study 1 examined the influence of different sentence contexts on word production in healthy younger and older adults. We compared matched contexts that predicted upcoming words, neutral contexts that did not predict a specific word, and mismatched contexts that predicted alternative words, as well as naming words without context. The likeliness ratings confirmed that targets were indeed most likely to follow questions in the matched contexts and least likely to follow questions in the mismatched contexts. As expected, older adults were slower overall in the naming task than younger adults. Participants in both age groups showed a (similar) significant Match effect, reflecting faster naming in the matched than neutral contexts. No Mismatch effect (difference between mismatch and neutral contexts) was found in either age group. Some effect of context was observed in both age groups, although this was only present when participants had to name pictures without context first.

2.4.1. Facilitatory effects of semantically matching contexts

Similar to Shao and Rommers' study with younger adults (2020), our data showed that target-word production was faster after matched than after neutral sentences. The Match effect is likely related to semantic priming effects, wherein words in the context prime related words, making their retrieval faster. Listeners might furthermore be predicting upcoming words and when these predictions are in line with the required response, production might be facilitated. This facilitation was found for both younger and healthy older adults to a similar degree. Previous research assessing context effects has mostly focused on comprehension and has returned mixed findings when comparing older and younger adults. Some of those studies have suggested similar benefits for older and younger adults (e.g., Kliegl et al., 2004) while others have suggested older adults benefit more (e.g., Pichora-Fuller et al., 1995) or less (e.g., Wlotko et al., 2012) from semantically related context than younger adults. However, many of these studies compared matched to incongruent or mismatched conditions, rather than to a neutral baseline. This often makes it difficult to attribute any context effects and age group similarities or differences to benefits associated with *matched* semantic content specifically. Our study shows that older adults can indeed benefit from semantic information matching upcoming target-word production. In this specific case, they did not benefit more than younger adults (unlike e.g., Pichora-Fuller et al., 1995). However, such additional benefits for older adults might only arise when

DOI: 10.1177/17470218241309602

processing demands are high (e.g., when the task is presented in background noise, as done in Pichora-Fuller et al., 1995). Furthermore, more compensatory effects (exhibiting a larger Match effect in older adults) might be more likely to arise in low-frequency target words with poorer lexical-phonological connections (James & MacKay, 2001). In contrast, our study used relatively high-frequency words, a familiarisation phase, and each word was produced multiple times to ensure the same words were used in all conditions. As a result, older adults may not have had to rely more heavily on semantic context than younger adults. However, the benefits observed are likely to require older adults to have intact semantic knowledge to facilitate priming of related words. Study 2 therefore firstly examined semantic knowledge in both younger and older adults, to test the hypothesis that older adults indeed continue to benefit from a semantic knowledge and vocabulary size at least comparable to that of younger adults. We furthermore examined whether individual differences in the size of context effects during language production can be explained by individual differences in semantic knowledge.

2.4.2. Mismatched sentence contexts

Picture-word interference paradigms without context have shown that distractor words semantically related to target pictures (e.g., target: ball, distractor word: frisbee) often slow down target word retrieval compared to unrelated distractors (e.g., hammer) in both younger and older adults (Taylor & Burke, 2002). This suggests that competing active semantic information can interfere with the production of intended words. However, in our study, mismatched sentence contexts did not influence target-word production times. Although we expected a Mismatch effect with slower responses when the target word was unexpected, these findings align with a previous comprehension study using comparable sentences (Haigh et al., 2022), where no Mismatch effects were found either. Similarly, Luke and Christianson (2016) showed no negative impact of unpredictable information in an eye-tracking study. Previous research (Brothers & Kuperberg, 2021) has suggested people engage less in predicting upcoming words when the context is uninformative. However, considering the observed Match effect, it is unlikely that the absent Mismatch effect in our study is due to the context's overall low informative level resulting in people stopping predicting upcoming information entirely.

While our mismatched targets were *unlikely* answers to the preceding question, they were not *impossible*. Indeed, while likeliness ratings differed significantly between mismatched and neutral sentences, the differences were small. This is likely related to the mismatch trials not being unexpected enough, although the difference can also be increased by making the neutral trials more neutral/predictable. Previous comprehension studies have found the largest processing costs when sentence endings are semantic anomalies (e.g., Federmeier & Kutas, 1999). The mismatched contexts in the current study may not have increased competition between word candidates sufficiently to

DOI: 10.1177/17470218241309602

incur a cost, in particular when comparing these effects to a neutral baseline (as opposed to a comparison to expected, matched contexts). Recent research also assessing production indeed showed only a small cost when unpredictable words were used that were closely related to the expected target (e.g., “skull” instead of “brain”) compared to a larger cost when the unpredictable word was not related to the expected target (Bannon et al., in press). Our exploratory analyses further suggested that, also within the mismatch context, responses were slower for target words that were less likely to occur in the sentence context. This suggests that a selection of impossible rather than unlikely target endings would perhaps be more likely to show significant interference effects. Variability between items in their (un)likeliness also likely contributes to the low split-half reliability observed for the Mismatch effect. However, age group did not interact with the effect of likeliness in the continuous rating analyses either, suggesting the older adults in this study were truly not affected more by target-word likeliness.

2.4.3. Neutral sentence contexts and ageing

Picture naming was faster following neutral contexts compared to no context, suggesting sentence context facilitated production. Furthermore, this context facilitation appeared to be similar for both age groups, although some analyses suggested the effect was slightly smaller for older adults. Pictures were repeated throughout the task and participants were exposed to them beforehand through a picture familiarisation phase. It is possible that this facilitated older adults’ overall retrieval (compared to not having seen the picture or word beforehand) and that older adults benefit more from context when words are harder to retrieve (e.g., without previous exposure or when using lower-frequency words). Older and younger adults also experienced the workload in No-Context and Context conditions comparably, suggesting there was no age-group difference in perceived effort involved in naming with or without context.

However, follow-up analyses suggested the facilitatory effect of context in terms of naming times might not be entirely driven by effects of context as such. While the order of naming with or without context first was counterbalanced across participants, only participants who completed the naming task without context first showed faster naming within context. Given that pictures were repeated, it is possible that participants benefited from naming those pictures without context first and therefore showed faster naming times in the sentence contexts. This could be because previous naming primed the lexical form (allowing for faster retrieval in the context task when completed second) and/or because participants expected the same pictures to appear again. A true effect of context facilitation on production should occur even if the context part is completed first. However, context effects varied between participants and it is possible that other differences (e.g., individual differences in terms of

DOI: 10.1177/17470218241309602

working memory) contributed to the variability observed in context effects. This is assessed in more detail in Study 2.

3. Study 2

3.1. Introduction

In Study 2, we aimed to further examine the potential mechanisms underlying the context effects. Specifically, we studied how age groups differed in terms of cognitive and language variables (such as semantic knowledge and control, inhibition, and working memory), and assessed how these variables contributed to the context effects observed in Study 1. Note that this study was pre-registered before completing Study 1 and the hypotheses therefore do not take into consideration the Study 1 findings (e.g., the absence of a Mismatch effect). However, Study 2 focused mostly on trying to explain *individual differences* in these context effects, which were indeed observed in Study 1, despite the absence of group-level differences.

3.1.1. Semantic knowledge

As introduced in Study 1, semantic knowledge has been found to increase with age (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003). For example, in a synonym selection task, Hoffman (2018) and Hoffman et al. (2018) asked younger and older adults to select the synonyms of probe words (e.g., *"which means the same as bombastic?"* answer: *pompous*, other answer options: *destructive, anxious, bickering*). In both studies, older adults provided significantly more correct answers than younger adults. Larger semantic knowledge scores might also relate to lexical retrieval. On the one hand, faster language production has been observed for participants with a larger vocabulary (Shao et al., 2014), suggesting greater semantic knowledge is associated with faster word retrieval. In older adults, their larger vocabulary may act as a compensatory defence against age-related lexical access difficulties (Juncos-Rabadán et al., 2010). On the other hand, having access to more words could create additional interference and disrupt language production (Ramscar et al., 2014). For example, Hoffman and colleagues (2018) found a negative relationship between semantic knowledge and coherence in connected speech.

The size of individuals' semantic knowledge stores might particularly relate to the Match effect measured in Study 1. The Match effect is expected to depend on participants having access to semantic knowledge and connections, which allow for semantic priming and predictions about upcoming, semantically related words. In Study 2, we therefore used a synonym judgement task to assess semantic knowledge in the younger and older adults tested in Study 1, as well as a potential relationship between semantic knowledge and context effects (in particular the Match effect). As

DOI: 10.1177/17470218241309602

discussed above, such relationship could go in two directions. Greater semantic knowledge and a larger vocabulary could facilitate word retrieval in matched contexts if they increase priming and allow participants to more easily predict upcoming words. Alternatively, having access to more semantic knowledge can increase interference and disrupt production (Ramscar et al., 2014). This could then result in smaller Match and potentially larger Mismatch effects.

3.1.2. Fluency

We also assessed semantic knowledge and lexical retrieval through verbal fluency tasks, which require participants to produce as many words as possible belonging to a specific category (semantic fluency) or beginning with a specific letter (letter fluency), within a specified time. Fluency has been linked to both lexical retrieval efficiency and semantic knowledge. Counter-intuitively, older adults often perform more poorly than younger adults on semantic fluency tasks while age-group differences tend to be smaller or absent on letter fluency tasks (e.g., Gordon et al., 2018; Kavé & Knafo-Noam, 2015). Here, we were predominantly interested in verbal fluency in general, given its links to both lexical retrieval efficiency and semantic knowledge across semantic and letter tasks (cf. Gordon et al., 2018). Verbal fluency has been associated with language production across younger and older adults (Higby et al., 2019). Greater semantic knowledge and lexical retrieval efficiency, as measured through verbal fluency scores, could help participants to benefit more from semantic connections in matched sentence contexts. We therefore examined whether verbal fluency (across letter and semantic fluency tasks) influenced the Match effect.

3.1.3. Semantic and domain general control

As discussed in the General Introduction, older adults have also shown declines in inhibition and semantic control. (e.g., Hasher & Zacks, 1988; Hoffman, 2018). These control mechanisms are believed to relate to language production, where speakers need to inhibit competitors in favour of producing required words (e.g., Faroqi-Shah et al., 2014). This could be particularly pertinent in scenarios wherein speakers are required to produce unexpected words (i.e., the mismatched sentences in Study 1). Hoffman (2018) used a global and feature semantic association task to assess semantic control. For instance, in the feature association task, participants selected the feature associate of a probe word (e.g., the item related in colour or size). In the low control manipulation (congruent trials), probe and target shared a semantic relationship *(e.g., "which is the same colour as cloud*", target: *snow*) and the distractors were not semantically related to the probe (e.g., *egg, step, basket*). In the high control manipulation (incongruent trials), the probe and target did not share a semantic relationship, but one of the distractors was semantically related to the probe (e.g., probe = *salt*, colour associate = *dove*,

DOI: 10.1177/17470218241309602

distractor = *pepper*). Older adults displayed poorer accuracy and longer RTs on the semantically incongruent trials, compared to younger adults (Hoffman et al., 2018, Hoffman, 2018).

Hoffman and colleagues also found semantic control to be related to coherence during language production, with participants who performed poorly on the semantic control task also producing less coherent speech. In Study 2, we therefore examined potential relationships between semantic control and context effects (in particular, the Mismatch effect). We expected participants who exhibited poorer performance on measures of semantic control to exhibit a larger (negative) Mismatch effect (greater interference from competing semantic information).

3.1.4. Inhibitory control

Domain general inhibitory control has also been found to decline with age. For example, older adults have been found to experience greater interference costs than younger adults from incongruent stimuli in colour Stroop tasks, where participants are asked to produce a response based on the text colour of presented words while ignoring word meaning (e.g., when the word “red” is presented in green text; Spieler & Faust, 1996; West & Alain, 2000). Similar age effects have been found on other inhibition tasks too (e.g., Hoffman, 2018) although they might depend on the task used (e.g., de Bruin & Della Sala, 2018; Rey-Mermet et al., 2018).

Similar to semantic specific control processes, domain general inhibitory control processes may also influence language production by facilitating the suppression of competing words. This could be especially relevant for the Mismatch effect, where participants have to produce an unexpected word while controlling interference from an expected word. Indeed, poorer inhibition skills can predict the level of coherence in individuals’ speech (Arbuckle and Gold, 1993; Yin & Peng, 2016). In contrast, however, Higby et al. (2019) did not find a relationship between inhibition and picture or object naming response times. In addition to semantic control, Study 2 therefore also assessed inhibition to examine whether poorer inhibition skills are associated with a larger Mismatch effect.

3.1.5. Short-term memory capacity

In line with resource theories, short-term memory capacity has been found to decline with age (Van der Linden et al., 1994; Waters & Caplan, 2003). For example, the mean number of items recalled in span tasks (including forward and backward digit, and letter and word span tasks) is lower in older adults (see Bopp & Verhaeghen, 2005, for a review). Short-term memory has also been associated with language processing (e.g., Feier & Gerstman, 1980; Kemper et al., 1989; Walsh & Baldwin, 1977). In Study 2, we therefore explored how age-related short-term memory declines (measured through the digit span task) may contribute to the Context effects (compared to no context). When producing

DOI: 10.1177/17470218241309602

words in context, speakers must hold the context within their working memory whilst planning and producing their verbal response. Producing words without context might not place the same demands on working memory resources. Thus, poorer working memory capacity might modulate how much participants can benefit from producing words in context and therefore the context effect in Study 1.

3.1.6. Lifestyle and social network

Finally, we included participants' education, lifestyle, and social network in the analysis in Study 2. Education may serve as a protective factor that may reduce word-retrieval difficulties associated with ageing (Gordon & Kindred, 2011). Furthermore, lifestyle activities have been associated with reduced cognitive changes with age (Scarmeas et al., 2003). We furthermore assessed the frequency and nature of social interactions, as previous research suggests that language difficulties can be associated with reduced social interactions (Burke & Shafto, 2004; Farrell et al., 2014). Based on the described relationships between educational, lifestyle, and social factors and language, in Study 2 we were interested in exploring whether these factors also contribute to the context effects in Study 1.

3.1.7. Rationale Study 2

In Study 2, we assessed potential differences between younger and older adults on measures of language and cognitive functioning (including semantic knowledge, fluency, semantic control, inhibition, and verbal short-term memory capacity), as well as in terms of lifestyle and social interactions. We also studied how these variables related to language production in the different contexts assessed in Study 1. For the match effect, we were particularly interested in semantic knowledge and fluency. For the mismatch context, although neither the younger nor older adult group exhibited a mismatch effect in Study 1, we were interested in whether individual differences were linked to semantic control and/or inhibitory control. Finally, we assessed whether the magnitude of the context effect (neutral context versus naming in isolation) was related to individuals' short-term memory capacity.

3.2. Methods

This study was pre-registered on the Open Science Framework: https://osf.io/wf8cm. The data are available on https://osf.io/8qexr/, together with Study 1.

3.2.1. Participants

Ethical approval was obtained from the Department of Psychology at the University of York. All participants who completed Study 1 were invited again to complete Study 2, with 45 older and 38

DOI: 10.1177/17470218241309602

younger adults taking part. Compared to the full set of participants in Study 1, the participant profile was comparable in terms of age (*M* younger adults = 24.61 years, *SD* = 4.79; *M* older adults = 69.09, *SD* = 4.12) and education (*M* younger adults = 16.57 years, *SD* = 2.99; *M* older adults = 15.60, *SD* = 3.36). The mean interval between completing Study 1 and 2 was 22.86 days (range = 6-45 days, *SD* = 7.65).

3.2.2. Materials/Tasks

Semantic knowledge

Semantic knowledge was assessed through a synonym judgement task (adapted from Wu & Hoffman, 2022). Participants viewed 67 word pairs and decided if the two words shared a similar meaning by pressing a keyboard button (S = related; D = different). Half of the word pairs were unrelated *(e.g., "formidable, "obdurate")* and the other half were related *("recondite", "abstruse").* There was no time limit per trial and the next trial started automatically when participants completed the previous trial. For each participant, we calculated an ISDT score (an index based on signal detection theory) to correct for guessing and response style (Huibregtse et al., 2002). This score takes into consideration participants' hit rate (proportion of correct "related" responses), as well as their false alarm rate (proportion of "related" responses to different pairs of words). This score ranges from 0-1 (a score closer to 1 indicates better performance, whereas a score closer to 0 indicates lower performance).

Verbal fluency

We measured both letter and semantic fluency. Participants completed three letter fluency trials (producing as many words as they could beginning with 'F', 'A', or 'S') and three semantic fluency trials ('animals', 'fruits', and 'items of clothing'). Each trial was 60s long. For each participant we computed a composite fluency score (the average of number of words produced across the letter and semantic trials). Fluency data from one older and one younger participant were excluded from the analyses because the recordings were of poor quality or empty.

Semantic Control

Semantic control was measured through two semantic association tasks (used in Hoffman, 2018). In the global association task, participants selected the word associated with a probe, from a set of possible answers. Within the low demand condition (50% of trials), there was a strong semantic relationship between the probe and target (e.g., probe*: town,* target*: city*). In the high demand condition, there was a weak semantic relationship between the probe and target (e.g., probe: *iron***,** target: *ring*).

DOI: 10.1177/17470218241309602

In the feature association task, participants selected the word matched with a probe word on a particular feature (e.g., colour or size, whilst ignoring distractor words). Half of the trials required participants to select the word most closely related in size, and the other half to select the word most closely related in colour to the probe. Half were congruent trials (target and probe shared a semantic relationship). For example, participants would see "which is the most similar in size to *door*?" with "*window*" being the target and none of the distractors *(bottle, report, factory)* sharing a semantic relationship with the probe. In the incongruent feature trials, the target and probe were not semantically related but one of the distractors was related to the probe. For example, participants would see *"*which is most similar in size to *ashtray?"*, with target: *diary* and distractor *cigarette*. There was no time limit (following Hoffman, 2018) and the next trial started automatically when they completed the previous trial.

In both tasks, half of the trials included four answer options, and the other half two; in the analyses we collapsed across number of options. In the main analysis, we calculated the accuracy cost within each task for each participant. This was the z-scored accuracy difference between the low and high control conditions in the global association task (global cost) and the low and high control conditions in the feature association task (feature cost). Additional analyses were conducted using RTs, for which we removed RTs two standard deviations above or below participants' conditional mean (following Hoffman, 2018, and considering this task did not have a time limit) as well as incorrect responses.

Inhibitory control

Inhibitory control was measured through a verbal and non-verbal Stroop task. In the verbal version, participants provided a keyboard response based on the colour of a written word. In congruent trials (n=28), word meaning and colour were the same (e.g., 'red' presented in red), while they differed in incongruent trials (n=28, e.g., the word 'blue' presented in red). Neutral trials used a non-colour word (n=28, e.g., 'flower' presented in blue). The non-verbal version was a digit-based task, where participants decided which of two presented numbers was larger physically in terms of text size whilst ignoring numerical size. In the congruent trials (n=28), the number which was physically bigger was also numerically bigger than the other number. In the incongruent trials (n=28), the numerically smaller number was presented in a larger font. In neutral trials (n=28), the same number was presented twice.

In each trial, participants saw a fixation cross for 500ms, followed by the stimulus. The next trial was presented as soon as a response was given or, if no response was provided, after 3000ms.

DOI: 10.1177/17470218241309602

A verbal and a non-verbal Stroop interference cost was computed for each participant as the mean RT difference between the neutral and incongruent trials in each task. By computing the difference between neutral and incongruent trials, we specifically looked at interference, leaving out the influence of facilitation on congruent trials. Prior to calculating Stroop cost, we first removed incorrect responses and RTs 2.5 *SD*s above or below the mean per condition and per participant. Participants' composite Stroop cost was calculated by computing the mean cost across the two tasks. The colour Stroop data from one older participant was not saved successfully and was therefore excluded from analysis.

Short-term memory span

Short-term memory was assessed using a digit span task. Participants viewed sequences of two to eight digits, with the sequence size increasing after every two consecutive trials (16 trials in total). Within each trial, participants viewed a fixation cross for 250ms, followed by the individual presentation of each digit. Digits were presented in the centre of the screen for 800ms before the screen automatically proceeded to the next digit. At the end of each trial, a text box appeared, and participants were asked to type all the digits they could remember from that trial. We calculated the proportion of correct sequences recalled by each participant. All participants saw all sequences, even if they made mistakes earlier on in the task (with shorter sequences).

Lifestyle and social network scores

Demographic details were derived from two questionnaires (administered in Study 1 and Study 2). Four participants did not provide the total number of years of formal education they had received. A lifestyle questionnaire (adapted from Scarmeas et al., 2003) asked participants to state for a range of activities (e.g., reading or traveling, 18 items) on a scale from 1 (never) to 3 (often) how often they did them during the year preceding the Covid-19 pandemic (the study was conducted during 2021, when many Covid-19 social distancing restrictions were still in place). Participants' total score on the questionnaire was computed as their 'lifestyle score' (min = 18, max score = 54). In addition, we assessed social network size by asking participants to estimate the number of people they had had regular contact with in the past 6 months (including face-to-face, by phone or mail, or on the Internet). This was assessed across categories including close friends, family, neighbours, co-workers, school/child relations, people who provide a service, and others. We added together the total number of people in each category to compute each participant's social network size (adapted from Bruine de Bruin & Parker, 2020). We removed 'social network size' from the analysis for participants who reported having been in regular contact with more than 1000 people in the past 6 months (*n*=3).

DOI: 10.1177/17470218241309602

3.2.3. Procedure

The experiment was conducted using Gorilla.sc (Anwyl-Irvine et al., 2020). Participants read a study information sheet and completed a consent form to confirm that they met the study criteria and agreed to participate. Participants also completed a sound check to ensure their microphone was working (required for the fluency task). After this, participants completed the tasks in the following order: synonym judgement, semantic control tasks, Stroop (verbal), Stroop (digit), digit span, letter and semantic fluency, and lifestyle/social questionnaire. The study took approximately 45-60 minutes.

3.2.4. Data analysis

First, we examined whether the age groups differed in their performance on these tasks, using independent t-tests. Next, we conducted three regression analyses assessing if the measures described above explained the match, mismatch, and context effects observed in Study 1. We z-scored the data for each continuous predictor variable (across age groups, given that part of our analysis aimed to examine individual differences in relation to language/cognitive abilities across age). Pearson's correlation coefficients between the predictors (all below .5) as well as VIF statistics (VIF values <2.5 for all predictors) suggested there were no multi-collinearity issues.

We conducted hierarchical multiple linear regression analyses with the Match, Mismatch, and Context effects as outcomes. In each model, demographic variables were inputted first (age, lifestyle score, gender, social network size, years of formal education); this was followed by the cognitive and language variables (synonym judgement ISDT score, fluency composite score, semantic control global cost, semantic control feature cost, Stroop composite interference cost, and digit span accuracy); and finally we also included interactions between the cognitive/language variables and age. Although all participants completed all tasks, there were some technical issues with some data files and we therefore removed missing data pairwise in the analyses. We computed split-half reliability estimates for the cognitive variables entered into the regression models. These were generally moderate (Spearman-Brown scores ranging between .43 and .74).

3.3. Results

3.3.1. Age group comparisons

Below, we first present the analyses comparing the age groups. In terms of semantic knowledge and verbal fluency, older adults outperformed the younger adults on the synonym judgement task (*M* older adults = .66, *SD* = .13, *M* younger adults = .47, *SD* = .16; $t(81) = 6.169$, $p < .001$, $d = 1.359$; see

DOI: 10.1177/17470218241309602

Supplementary Figure 1) and the verbal fluency task (*M* older adults = 18.37, *SD* = 3.35, *M* younger adults= 16.95, *SD* = 2.86; $t(79) = 2.037$, $p = .045$, d=.454, see Supplementary Figure 2).

In terms of semantic control (see Supplementary Figure 3), our pre-registration focused on accuracy scores. For both tasks, we computed a semantic control score by taking the difference between strong and weak trials (global association) and congruent and incongruent trials (feature association). In the global association task, older adults' semantic control cost (*M* = 4.35%, *SD* = 8.61) was, surprisingly, significantly *lower* than the younger adults' cost (*M* = 11.40%, *SD* = 6.76; $t(81) = -4.093$, $p < .001$, $d = -0.902$). In the feature association task, the numerical pattern went in the same direction but was not significant (older adults *M* = 12.78%, *SD* = 19.67; younger adults *M* = 17.11%, *SD* = 23.35; $t(81) = -0.917$, $p = .362$, $d = -0.202$). The RT data showed no significant cost difference in the global association task (older adults *M* = 836.24ms, *SD* = 478.02; younger adults *M* = 699.74ms, *SD* = 487.26; $t(81) = 1.285$, $p = .203$, $d = .283$). In the feature association task, the RT cost was significantly higher in the older (*M* = 1277.61ms, *SD* = 815.00) than in the younger age group (*M* = 703.84ms, *SD* = 735.63; $t(80) = 3.325$, $p = .001$, $d = 0.736$). Similar RT results were obtained when comparing the z-scored RTs, considering older adults responded more slowly overall. Given the potential speed-accuracy trade off (older adults showing larger RT costs with smaller accuracy costs), we also computed inverse efficiency scores (RT/percentage correct). These did not differ between age groups for the global task ($t(81) = -.267$, $p = .790$, $d = -0.059$) or for the feature task ($t(80) = -.497$, $p = .620$, $d = -0.110$).

The Stroop analysis, as pre-registered, focused on the RTs only. This analysis excluded incorrect responses. Accuracy in the non-verbal Stroop task was 82.07% (*SD* = 4.66) for older adults and 80.03% (*SD* = 9.19) for younger adults. In the verbal Stroop task, accuracy was 74.73% (*SD* = 13.97) for older adults and 81.21% (*SD* = 4.79) for younger adults. The Stroop RT interference cost was significantly larger in the older (*M* Stroop = 82.41ms, *SD* = 80.29) than younger adults (*M* = 39.54ms, *SD* = 85.18; $t(80) = 2.344$, $p = .022$, $d = 0.519$; see Supplementary Figure 4). This significant difference remained when analysing the z-scored RTs (considering overall slower responses in older adults).

Short-term memory capacity (measured through the digit span task) did not differ between older adults (*M* = 65.56%, *SD* = 14.06) and younger adults (*M* = 65.41%, *SD* = 17.30; $t(81) = 0.041$, $p = .967$, $d = 0.009$; see Supplementary Figure 5).

Finally, social network size was slightly larger in the younger participants (*M* = 33.03, *SD* = 23.09) compared to the older participants (*M* = 27.73, *SD* = 17.25) but this difference was not significant ($t(78) = -1.174$, $p = .244$, $d = -.264$). Older adults scored significantly higher on the lifestyle questionnaire (*M* = 37.64, *SD* = 3.93) in comparison to the younger participants (*M* = 34.66, *SD* = 3.84; $t(81) = 3.483$, $p < .001$, $d = .767$).

DOI: 10.1177/17470218241309602

3.3.2. Hierarchical Regression

Hierarchical linear regressions were computed next to measure the contribution of each predictor on each context effect from Study 1 (Match, Mismatch, and Context effect). Supplementary figures 6 to 8 show the relationships between the synonym judgement, verbal fluency, the two semantic control (global and feature association), inhibition, and digit span tasks with the Match, Mismatch, and Context effects.

Match effect

None of the included cognitive variables were significant individual predictors of the Match effect (see Table 7). The contribution of demographic variables together (Model 1) was close to significance, with social network reaching significance. This suggests that people with a larger social network showed a smaller Match effect. Models 2 and 3, which included the cognitive and language variables, and age interactions with those cognitive and language variables, did not contribute significantly beyond the model only including demographic variables.

--- Insert Table 7 about here ---

Mismatch effect

The mismatch effect was not explained significantly by any of the individual predictors. None of the three models reached significance either (see Table 8)

--- Insert Table 8 about here ---

Context effect

Finally, none of the included variables were significant predictors of the Context effect (see Table 9). None of the three overall models reached significance either.[1]

[1] Considering that (negative) age effects on the semantic control task were observed in terms of RTs but not accuracy, we re-ran the regression analyses with the RT global and feature association costs. These analyses did not find significant effects of the included cognitive predictors on Matched, Mismatched, or Context effects either, with the exception of Age x Global RT cost being a predictor of the Mismatch effect. For younger adults, a larger semantic-control cost was associated with a slightly smaller Mismatch effect while for older adults a larger semantic-control cost was associated with a slightly larger Mismatch effect. However, this control cost was not a significant predictor for either age group in separate regression analyses per group. Finally, we checked whether any of the cognitive and language scores related significantly to overall naming times (as

DOI: 10.1177/17470218241309602

3.4. Discussion

In Study 2, we assessed how the older and younger adults tested in Study 1 differed on various cognitive abilities including semantic knowledge and control, inhibition, and short-term working memory capacity. Semantic knowledge and fluency scores were larger for older than younger adults. However, older adults performed more poorly in terms of inhibition costs and (some aspects of) semantic control, although they did outperform younger adults in terms of semantic control accuracy. The language and cognitive variables assessed did not relate to language-production match, mismatch, or context effects tested in Study 1. This will be discussed further in the General Discussion.

3.4.1. Semantic knowledge

Corroborating previous research (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003), older adults performed significantly better on the synonym judgement task than the younger adults, suggesting that semantic knowledge was higher in the older than younger adults. Furthermore, composite verbal fluency was also higher in the older adult group. In line with previous research (e.g., Gordon et al., 2018), exploratory analyses presented in the Supplementary Materials showed this benefit for older adults was driven by the letter fluency trials rather than the semantic fluency trials. This aligns with a frequently, although seemingly paradoxically, observed pattern in the literature reflecting older adults are more likely to experience difficulties on the semantic than letter fluency task (cf. Gordon et al., 2018). Although letter fluency is often associated with executive control and would therefore be expected to be influenced more strongly by age, previous research has suggested letter fluency relies more heavily on vocabulary knowledge. The finding that our older adults outperformed the younger adults on this fluency task specifically supports, in line with the synonym judgement task and the literature, the interpretation that older adults continue to benefit from their (larger) vocabulary knowledge. In contrast, semantic fluency has been found to be more heavily influenced by lexical retrieval speed (Gordon et al., 2018). Indeed, our supplementary analyses showed no significant age-group difference here. If anything, older adults performed a little worse than younger adults on this task.

3.4.2. Semantic and domain general control

To measure semantic control, we used global association and feature association tasks (Hoffman, 2018). Both showed costs (poorer accuracy and longer RTs) in the conditions associated with higher

---

opposed to the context effects specifically). Analyses using RTs in the no-context task showed no relationship with any of the predictors either.

DOI: 10.1177/17470218241309602

control demands. In terms of accuracy, both measures showed higher costs in younger than older adults, although this difference was not statistically significant in the feature association task. These accuracy effects were contrary to our predictions regarding older adults experiencing difficulties with semantic control (and the findings observed in Hoffman, 2018). Some previous research has reported age-related increases in motivation and engagement with lab-based tasks (Frank et al. 2015; Jackson & Balota, 2012). This could explain why older adults were less hindered by the more challenging conditions in the semantic control tasks, especially given that there was no time constraint within these tasks. On the other hand, older adults did exhibit a greater RT cost, although only on the feature association measure. The feature association task was the task showing the largest accuracy and RT costs across age groups, suggesting older adults showed larger RT costs only on the more demanding control task. The combination of accuracy and RT findings suggests the older adults needed more time during the semantic-control task to suppress irrelevant features, but were able to achieve higher accuracy by doing this. Older adults also showed a larger interference Stroop cost. This was especially the case in the verbal Stroop task, where older adults also showed lower accuracy than younger adults (suggesting this larger RT cost was not due to a speed-accuracy trade off). These findings suggest older adults showed poorer semantic (on some measures) and inhibitory control in terms of response times, lending support to the inhibition deficit hypothesis (cf. also Hoffman, 2018).

3.4.3. Working memory

Finally, working memory capacity did not differ between the two age groups. Forward digit span tasks measuring storage capacity might not be sufficiently sensitive to age-related declines in working memory, relative to tasks that focus on both storage and information manipulation such as the backward digit span task (e.g., Babcock & Salthouse, 1990; Bopp & Verhaeghen, 2005). Furthermore, to adjust the task to an online environment, our participants were able to continue onto longer sequences even if they made errors in shorter trials, which may have influenced performance on this task relative to traditional task versions not allowing this.

3.4.4. Lifestyle and social network

Older and younger adults showed some differences in terms of their lifestyle and social network. Older adults reported higher engagement in the lifestyle activities we assessed than the younger adults, although this may partly be related to the type of activities assessed (e.g., gardening). Social network size did not differ significantly between the age groups.

DOI: 10.1177/17470218241309602

4. General discussion and conclusion

Together, Study 1 and 2 show that both younger and older adults' language production is influenced by the sentence context preceding a target word. Semantically matching contexts facilitated both younger and older adults' lexical retrieval. These context effects can be explained through more automatic mechanisms such as semantically related words being primed by previously presented words, therefore facilitating lexical retrieval during production. In this specific task context, where participants had to alternate between comprehending speech and producing a response, this priming operated cross-modally. Linked to priming, speakers can use the speech presented to them to predict suitable upcoming responses to their conversation partner. Our study shows that both younger and older adults continue to use sentence context to facilitate their own responses, with no significant difference between the groups. This suggests older adults may continue to use their semantic knowledge to facilitate lexical retrieval during production.

These findings support theories arguing that semantic knowledge is more developed in later adulthood, and that this can benefit language production processes (Burke et al., 1991). The Transmission Deficit Hypothesis argues that older adults' slower lexical retrieval is related to slower connections between the lexical and phonological level, with the semantic level staying intact. Our findings, in particular regarding the Match effect, support this hypothesis and suggest the preservation or proliferation of representations within the semantic system can facilitate connections between the semantic and lexical level. Our study did also reveal an age-group difference in terms of overall RTs, suggesting (in line with the Transmission Deficit Hypothesis and the general literature) that overall production was slower in older adults. This could be related to older adults showing slower lexical retrieval, although without specifically examining lexical-phonological retrieval difficulties in the current study, it might also reflect general processing speed.

Although individual differences in participants' semantic knowledge and fluency were not directly related to the size of the Match effect (Study 2), it is possible that a certain degree of semantic knowledge is a prerequisite for speakers to benefit from matching contexts. With our older adults on average outperforming the younger adults on semantic knowledge and fluency tests, this prerequisite seemed to have been met by the older adults as a group and almost all older adults at the individual level. Indeed, only four of the older adults' semantic knowledge scores fell below the mean score for the younger adults. However, in the absence of a direct relationship between semantic knowledge and the Match effect, the exact contribution of semantic knowledge in older adults requires further research.

The finding that older adults showed the same Match (and no Mismatch) effect as younger adults while showing more semantic knowledge argues against previous research suggesting

DOI: 10.1177/17470218241309602

increased (semantic) knowledge in older adults could potentially create more interference (e.g., Ramscar et al., 2014, cf. Hoffman et al., 2018, showing adults with greater semantic knowledge were less coherent in connected speech). If larger semantic knowledge is associated with greater interference, our older adults should have shown a smaller Match (and potentially larger Mismatch) effect than younger adults at the group level.

Finally, it is worth further exploring the role of lifestyle variables in relation to semantic connections in sentence contexts. Social network size predicted the Match effect during language production. People with a smaller social network showed greater facilitation from matched contexts. This relationship was small and not significant in all models and does therefore require further research. It is, however, in line with previous research suggesting a relationship between social interactions and language difficulties (Farrell et al., 2014).

Contrary to our expectations, neither age group showed an interference cost during mismatching contexts (that were designed to prime a word other than the target). Furthermore, neither semantic nor inhibition costs (measured in Study 2) were significant predictors of the Mismatch effect in the main analysis. Although some previous studies have shown a relationship between semantic and inhibitory control and certain aspects of language production (e.g., Hoffman et al., 2018), our findings align with previous research showing no direct relationship between inhibition and picture naming times in younger and older adults (Higby et al., 2021). This could suggest that measures related to the types of words older adults use (e.g., speech coherence, Hoffman et al., 2018) could be more closely related to one's ability to suppress interfering information than speed of lexical retrieval (as examined in Higby et al., 2021). We did also expect such a relationship (with semantic and inhibitory control) to arise in our mismatch contexts, which specifically required the production of an unexpected rather than expected target word. With this Mismatch effect not arising at the group level in either age group, however, it is very likely that a stronger mismatching sentence context is necessary for any role of semantic or inhibitory control abilities to emerge.

Study 2 did show that older adults' responses were slower during high-control trials in both semantic and inhibitory control tasks, corroborating previous findings that these cognitive abilities can decline in old age (e.g., Hoffman, 2018; Spieler et al., 1996). This suggests that the absence of a Mismatch effect in the group of older adults was not the consequence of recruiting a sample of older adults with particularly high or fully preserved semantic or inhibitory control. Rather, it suggests that interference in the mismatched sentences might have been too weak to lead to a noticeable impact on language production. Future research will need to study the potential relationship between age-related changes in inhibitory or semantic control and stronger context violations during language production. Furthermore, such research might want to include older adults with more difficulties in

DOI: 10.1177/17470218241309602

terms of their semantic and inhibitory control. While age-group differences were observed in the current study, they were not present on all semantic control tasks and measures, and older adults might have slowed down their responses to achieve a higher accuracy level. As is common in these types of studies, the older adults (like the younger adults) had a relatively high level of education and a relatively active lifestyle. Research including a wider range of older adults from various backgrounds, including lower socio-economic status, would be more representative of the general population and might be more likely to capture age-related changes in terms of interference effects during cognitive and language-production tasks. Including a larger sample size would also be beneficial, as the current sample size might have limited the power to detect effects related to individual differences in Study 2. Furthermore, the older adults in our study were quite young (mean age <70 years). Future research might also want to include older adults with a higher age, including more participants over 75 years old. It is important to underline, however, that despite not observing the age-related changes in the context effects we investigated, in Study 1 we did observe the expected age-related slowing in older adults' overall naming RTs. This is in line with data that show that RTs are sensitive to age-related changes in language production from the age of fifty (Verhaegen & Poncelet, 2013).

Finally, previous research comparing age-group differences during language production with and without context has suggested any age effects might be less likely to be observed during connected speech in context (Kavé & Goral, 2017). However, such comparison across studies is made more difficult by the range of measures used (e.g., hesitations, naming times, circumlocutions). Furthermore, in connected speech, older adults can use compensatory mechanisms to mask lexical-retrieval difficulties. The current study therefore compared language production in older and younger adults in the form of picture naming with and without sentence context. This allowed us to compare the same measure (naming times) and removed any compensatory strategies that could be used during free speech. If older adults' lexical-retrieval difficulties are reduced during context, context effects (faster picture naming in a sentence context) should have been larger for older than younger adults. That was not the case; if anything, some analyses suggested the context effect was smaller in older adults. In combination with older adults not benefiting more from matched sentence contexts, this suggests that previously observed lexical-retrieval difficulties in older adults in picture-naming tasks are not purely due to somewhat artificial tasks requiring production of words without any context.

While the current study included producing words in response to a preceding question (as is common behaviour when people respond to other speakers), we did not examine a speaker's lexical retrieval within their own connected speech. Therefore, future research is needed to better understand the various mechanisms older adults might use when retrieving words in connected

DOI: 10.1177/17470218241309602

speech during different types of contexts. Context effects in the current study were not related to any of the language or cognitive abilities tested, including digit-span performance. This suggests that keeping the question in mind in a context as compared to naming pictures in isolation was not modulated by working memory capacity as assessed through this task. Sentences in our study were relatively short and simple. It is possible that more complex sentences, or producing words in connected speech and interactions with others, do tax short-term memory capacity more strongly (e.g., Kemper, 1986).

In conclusion, our findings suggest that the contexts within which words are produced have significant consequences for lexical-retrieval efficiency. Our study provides an important first step in comparing word production with and without context, while using the same measure (naming response times) to bridge the gap between existing literatures. Within the context of cognitive ageing, our findings highlight the important role of preserved semantic networks within older adults who can continue to benefit from context when retrieving words to respond to questions asked by their conversation partner.

Supplementary Material

The Supplementary Material is available at: qjep.sagepub.com

Data Accessibility Statement

The data from the present experiment are publicly available at the Open Science Framework website: https://osf.io/8qexr/

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388-407.

Arbuckle, T. Y., & Gold, D. P. (1993). Aging, inhibition, and verbosity. *Journal of Gerontology*, *48*(5), 225-232.

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(*1), 65–77.

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging*, *5*(3), 421-428.

Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999). Cross-modal semantic and

DOI: 10.1177/17470218241309602

homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society*, *5*(7), 626-640.

Bannon, J., Gollan, T. H., & Ferreira, V. S. (in press). Is predicting during language processing worth it? Effects of cloze probability and semantic similarity on failed predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Barresi, B. A., Nicholas, M., Tabor Connor, L., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition, 7*(3), 169-178.

Boersma, P. & Weenink, D. (2022). *Praat: doing phonetics by computer [computer program]*. http://www.praat.org./

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), 223-233.

Boudiaf, N., Laboissière, R., Cousin, E., Fournet, N., Krainik, A., & Baciu, M. (2018). Behavioral evidence for a differential modulation of semantic processing and lexical production by aging: a full linear mixed-effects modeling approach. *Aging, Neuropsychology, and Cognition*, *25*(1), 1-22.

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language*, *116*, 104174.

Bruine de Bruin, W., Parker, A. M., & Strough, J. (2020). Age differences in reported social networks and well-being. *Psychology and Aging*, *35*(2), 159-168.

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.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science, 13*(1), 21-24.

Carrol, G. (2023). Old Dogs and New Tricks: Assessing Idiom Knowledge Amongst Native Speakers of Different Ages. *Journal of Psycholinguistic Research*, *52*(6), 2287-2302.

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

Connor, L. T., Spiro, A. III, Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during

DOI: 10.1177/17470218241309602

adulthood. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 59*(5), 203–209.

De Bruin, A., & Sala, S. D. (2018). Effects of age on inhibitory control are affected by task-specific features. *Quarterly Journal of Experimental Psychology*, *71*(5), 1219-1233.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283-321,

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808-816.

Faroqi-Shah, Y., Sampson, M., Baughman, S., & Pranger, M. (2014). Inhibitory control, word retrieval and bilingual aphasia: Is there a relationship?. Frontiers in Psychology.

Farrell, M. T., Zahodne, L. B., Stern, Y., Dorrejo, J., Yeung, P., & Cosentino, S. (2014). Subjective word-finding difficulty reduces engagement in social leisure activities in Alzheimer's disease. *Journal of the American Geriatrics Society*, *62*(6), 1056-1063.

Faust, M. E., Balota, D. A., & Multhaup, K. S. (2004). Phonological blocking during picture naming in dementia of the Alzheimer type. *Neuropsychology*, *18*(3), 526-536.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149-161.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, Words, Sentences: Age-Related Changes Across Levels of Language Processing. *Psychology and Aging, 18*(4), 858–872.

Feier, C. D., & Gerstman, L. J. (1980). Sentence comprehension abilities throughout the adult life span. *Journal of Gerontology*, *35*(5), 722-728.

Ferré, P., Jarret, J., Brambati, S., Bellec, P., & Joanette, Y.(2020). Functional Connectivity of Successful Picture-Naming: Age-Specific Organization and the Effect of Engaging in Stimulating Activities. *Frontiers in Aging Neuroscience, 12*(535770), 1-14.

Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults’

DOI: 10.1177/17470218241309602

reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, *30*(2), 266-278.

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and its breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics,* pp. 350-374. Cambridge University Press.

Goral, M., Spiro III, A., Albert, M. L., Obler, L. K., & Connor, L. T. (2007). Change in lexical retrieval skills in adulthood. *The Mental Lexicon, 2*(2), 215-238.

Gordon, J. K., & Kindred, N. K. (2011). Word retrieval in ageing: An exploration of the task constraint hypothesis. *Aphasiology*, *25*(6-7), 774-788.

Gordon, J. K., Young, M., & Garcia, C. (2018). Why do older adults have difficulty with semantic fluency?. *Aging, Neuropsychology, and Cognition*, *25*(6), 803-828.

Grange, J.A. (2015). trimr: An implementation of common response time trimming methods. R package version 1.0.1. https://cran.r-project.org/web/packages/trimr/index.html

Haigh, P., Hanif, N., & de Bruin, A. (2022). Diving into a pool or volcano? Examining the influence of sentence context and task demands on sentence reading in younger and older adults. *Plos one*, *17*(12), e0279555.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

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: Advances in research and theory,* Vol. 22, pp. 193–225). Academic Press.

Heller, R. B., & Dobbs, A. R. (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and Aging*, *8*(3), 443-450.

Higby, E., Cahana-Amitay, D., Vogel-Eyny, A., Spiro III, A., Albert, M. L., & Obler, L. K. (2019). The role of executive functions in object-and action-naming among older adults. *Experimental Aging Research*, *45*(4), 306-330.

Hoffman, P. (2018). An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. *Scientific Reports, 8*(1), 8145.

Hoffman, P., Loginova, E., & Russell, A. (2018). Poor coherence in older people's speech is explained by impaired semantic and executive processes. *eLIFE, 7*(e38907).

Hoyau, E., Boudiaf, N., Cousin, E., Pichat, C., Fournet, N., Krainik, A., Jaillard, A., & Baciu, M. (2017). Aging Modulates the Hemispheric Specialization during Word Production. Frontiers *in Aging Neuroscience, 9*(125).

DOI: 10.1177/17470218241309602

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing*, *19*(3), 227-245.

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: converging evidence from the Sustained Attention to Response Task and reading for comprehension. *Psychology and Aging*, *27*(1), 106-119.

James, L. E., & MacKay, D. G. (2001). HM, word knowledge, and aging: upport for a new theory of long-term retrograde amnesia. *Psychological Science*, *12*(6), 485-492.

JASP Team (2024). JASP (Version 0.18.3)[Computer software].

Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*(3), 611-625.

Juncos-Rabadán, O., Facal, D., Rodríguez, M. S., & Pereiro, A. X. (2010). Lexical knowledge and lexical retrieval in ageing: Insights from a tip-of-the-tongue (TOT) study. *Language and Cognitive Processes*, *25*(10), 1301-1334.

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging, Neuropsychology, and Cognition, 24*(5), 508-527.

Kavé, G., & Halamish, V. (2015). Doubly blessed: older adults know more vocabulary and know better what they know. *Psychology and Aging, 30*(1), 68-73.

Kavé, G., & Knafo-Noam, A. (2015). Lifespan development of phonemic and semantic fluency: Universal increase, differential decrease. *Journal of clinical and Experimental Neuropsychology*, *37*(7), 751-763.

Kavé, G., & Yafé, R. (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: Independence or interdependence of skills?. *American Journal of Speech-Language Pathology, 23*(1), 36-45.

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics*, *7*(3), 277-287.

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

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging*, *12*(2), 362-371.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262-284.

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging*, *8*(1), 34-43.

DOI: 10.1177/17470218241309602

Lorenz, A., Zwitserlood, P., Regel, S., & Abdel Rahman, R. (2019). Age-related effects in compound production: Evidence from a double-object picture naming task. *Quarterly Journal of Experimental Psychology*, *72*(7), 1667-1681.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60.

MacKay, A., Connor, L. T., & Storandt, M. (2005). Dementia does not explain correlation between age and scores on Boston Naming Test. *Archives of Clinical Neuropsychology, 20*(1), 129-133.

Murphy, D. R., Craik, F. I., Li, K. Z., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging*, *15*(2), 323-334.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in healthy aging. *Cortex, 21*(4), 595-606.

ONS, (2021). Profile of the older population living in England and Wales in 2021 and changes since 2011. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/profileoftheolderpopulationlivinginenglandandwalesin2021andchangessince2011/2023-04-03#:~:text=Census%202021%20results%20show%20the,from%2016.4%25%20to%2018.6%25.

Ossher, L., Flegal, K. E., & Lustig, C. (2013). Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition, 20*(2), 220-242.

Parsons, S. (2021). Splithalf: robust estimates of split half reliability. *Journal of Open Source Software, 6*(60), 3041.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593-608.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859-862.

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*(1), 5-42.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448-465

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition?

DOI: 10.1177/17470218241309602

Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501-526.

Saling, L. L., Laroo, N., & Saling, M. M. (2012). When more is less: Failure to compress discourse with re-telling in normal ageing. *Acta Psychologica, 139*(1), 220–224.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Scarmeas, N., Zarahn, E., Anderson, K. E., Habeck, C. G., Hilton, J., Flynn, J., ... & Stern, Y. (2003). Association of life activities with cerebral blood flow in Alzheimer disease: implications for the cognitive reserve hypothesis. *Archives of Neurology*, *60*(3), 359-365.

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.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 89695.

Shao, Z., & Rommers, J. (2020). How a question context aids word production: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology, 73*(2), 165-173.

Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging*, *15*(2), 253-258.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 461-479.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and Aging*, *17*(4), 662-676.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332-339.

Van der Linden, M., Brédart, S., & Beerten, A. (1994). Age-related differences in updating working memory. *British Journal of Psychology*, *85*(1), 145-152.

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.

Walsh, D. A., & Baldwin, M. (1977). Age differences in integrated semantic memory. *Developmental Psychology*, *13*(5), 509-514.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory

DOI: 10.1177/17470218241309602

measures. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 550-564.

Wen, H., & Dong, Y. (2023). The effect of ageing on confrontation naming in healthy older adults: A three-level meta-analysis. *Journal of Cognitive Psychology, 35*(4), 480-508.

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*(2), 179-189

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975-988.

Wu, W., & Hoffman, P. (2022). Validated measures of semantic knowledge and semantic control: Normative data from young and older adults for more than 300 semantic judgements. *Royal Society Open Science*, *9*(2), 211056.

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*, 175744.

DOI: 10.1177/17470218241309602

Figure Captions

Figure 1. Box plots displaying untransformed mean RTs (ms) by age group (left panel: younger adults, right panel: older adults) and naming context. Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers.

Figure 2. Box plots showing the facilitatory effects of matched contexts (left), no Mismatch effect in either group (middle), and facilitatory effect of neutral (context) relative to naming without context (right) in both younger and older adults. Box plot height denotes interquartile range, vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent potential outliers.

DOI: 10.1177/17470218241309602

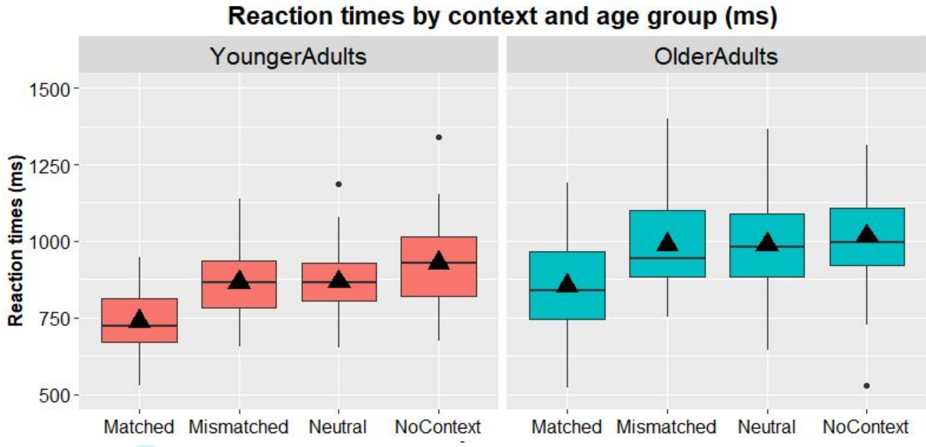

Figure 1. Box plots displaying untransformed mean RTs (ms) by age group (left panel: younger adults, right panel: older adults) and naming context. Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers.

DOI: 10.1177/17470218241309602

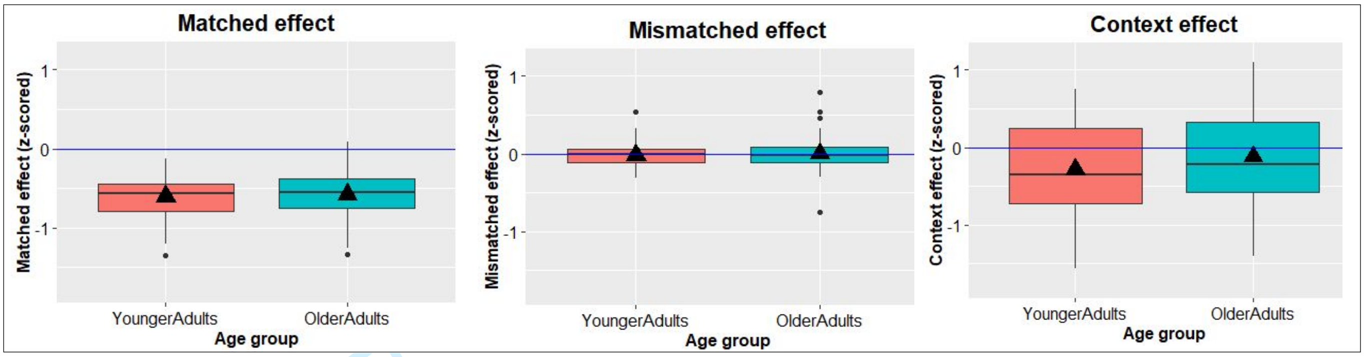

Figure 2. Box plots showing the facilitatory effects of matched contexts (left), no Mismatch effect in either group (middle), and facilitatory effect of neutral (context) relative to naming without context (right) in both younger and older adults. Box plot height denotes interquartile range, vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent potential outliers.

DOI: 10.1177/17470218241309602

Target picture	Matched	Mismatched	Neutral	No context
Egg	*What did the chef crack?*	*What did the baby shake?*	*What did the father ask his son to bring?*	No question
Rattle	*What did the baby shake?*	*What did the chef crack?*	*What did the father ask his son to bring?*	No question

Table 1. The first three columns relate to the three types of sentence contexts used in the study, including example sentences. The question always preceded the presentation of a target picture, which participants had to name. In the 'no context' condition, participants just named the picture (without hearing a question beforehand).

DOI: 10.1177/17470218241309602

	N	Age (in years)	Formal education (in years)	Graduate education	Sex		Handedness	
					Male	Female	Left	Right
Younger	48	24.25 (4.6) 18-35	16.5 (2.8)	29	25	23	10	38
Older	48	69 (4.1) 65-77	15.6 (3.4)	31	25	23	4	44

Table 2. Details of participants included in Study 1: age and formal education (mean number of years, and with standard deviations in parentheses, and age range below); graduate education (total number of participants who had completed at least an undergraduate degree); sex and handedness (total number of participants belonging to each category).

DOI: 10.1177/17470218241309602

	Younger adults	Older adults
Likeliness rating (1-5)		
Matched	4.89 (0.13)	4.88 (0.26)
Neutral	2.70 (0.87)	3.12 (0.83)
Mismatched	2.27 (0.79)	2.55 (0.91)

Table 3. Pilot 1 likeliness data: mean likeliness ratings for matched, neutral and mismatched question-answer pairs obtained from younger and older adults. Note that contrary to the main experiment, the pilot used a 1-5 rating scale (1 = very unlikely, 5 = very likely).

DOI: 10.1177/17470218241309602

	Younger adults	Older adults
Likeliness rating		
Matched	6.69 (0.30)	6.72 (0.28)
Neutral	3.63 (0.64)	3.30 (0.73)
Mismatched	3.02 (0.56)	2.89 (0.53)

Table 4. Likeliness ratings obtained from participants completing the full study. Note that contrary to the pilot, the main study used a 1-7 rating scale (1 = very unlikely, 7 = very likely).

DOI: 10.1177/17470218241309602

Context	Younger adults	Older adults
Without context	927.95 (141.75)	1016.28 (164.47)
Matched	738.31 (102.86)	852.66 (145.03)
Neutral	867.25 (106.57)	988.50 (139.05)
Mismatched	864.16 (105.66)	988.36 (150.16)

Table 5. Mean picture naming times (and standard deviations) per Age group and Context

DOI: 10.1177/17470218241309602

	Younger adults	Older adults
NASA		
Single	24.59 (14.65)	24.91 (14.61)
Context 1	27.29 (13.63)	26.70 (15.60)
Context 2	26.30 (15.77)	25.23 (15.69)
After Naming	25.02 (16.42)	23.70 (18.77)
Overall NASA	29.44 (17.92)	28.48 (21.91)

Table 6. The first four rows show the means and standard deviations of older and younger adults' NASA workload ratings after each naming block. 'Overall NASA' reflects the weighted after-naming score for each age group.

DOI: 10.1177/17470218241309602

model 1				model 2				model 3			
	B	*SE*	*P*		*B*	*SE*	*P*		*B*	*SE*	*P*
Intercept	-0.54	0.04	<.001	*Intercept*	-0.55	0.04	<.001	*Intercept*	-0.52	0.06	<.001
Age	-0.02	0.03	.632	*Age*	< - 0.001	0.04	.999	*Age*	0.02	0.05	.666
Education	-0.04	0.03	.164	*Education*	-0.04	0.03	.219	*Education*	-0.03	0.04	.332
Lifestyle	0.06	0.03	.059	*Lifestyle*	0.07	0.04	.057	*Lifestyle*	0.06	0.04	.111
Social net	-0.07	0.03	.035	*Social net*	-0.07	0.03	.034	*Social net*	-0.07	0.04	.068
Gender	-0.10	0.06	.106	*Gender*	-0.08	0.07	.226	*Gender*	-0.09	0.07	.175
				Synonym judgement	-0.001	0.04	.981	*Synonym judgement*	-0.02	0.05	.716
				Fluency	-0.04	0.04	.292	*Fluency*	-0.02	0.04	.556
				Global Cost	-0.01	0.04	.751	*Global Cost*	-0.03	0.04	.534
				Feature Cost	0.01	0.04	.860	*Feature Cost*	-0.01	0.04	.893
				Inhibition	-0.01	0.04	.864	*Inhibition*	-0.02	0.04	.546
				Digit span	0.01	0.03	.677	*Digit span*	0.003	0.03	.918
								Age Synonym judgement*	-0.07	0.05	.148
								Age Fluency*	-0.03	0.04	.459
								Age Global Cost*	0.07	0.05	.143
								Age Feature Cost*	-0.01	0.04	.756
								Age Inhibition*	-0.02	0.04	.496
								Age Digit span*	-0.02	0.03	.627
Model 1: (F(5,70) = 2.252, p = .059) Total variance explained: 13.9%				Model 2: (*FChange*(6,64) = .328, p = .920) Variance explained relative to model 1: 2.6%				Model 3: (*FChange*(6,58) = .850, p = .537) Variance explained relative to model 2: 6.8%			

Table 7. Hierarchical regression table showing the contributions of all predictors to the Match effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

DOI: 10.1177/17470218241309602

model 1				model 2				model 3			
	B	***SE***	***P***		***B***	***SE***	***P***		***B***	***SE***	***P***
Intercept	0.03	0.03	.281	*Intercept*	0.03	0.03	.393	*Intercept*	0.06	0.04	.102
Age	0.02	0.02	.495	*Age*	0.03	0.03	.319	*Age*	0.03	0.03	.413
Education	-0.02	0.02	.432	*Education*	-0.02	0.02	.349	*Education*	-0.03	0.02	.211
Lifestyle	0.01	0.02	.626	*Lifestyle*	0.01	0.02	.798	*Lifestyle*	0.003	0.03	.904
Social net	- 0.001	0.02	.964	*Social net*	-0.002	0.02	.919	*Social net*	0.01	0.02	.599
Gender	-0.08	0.04	.059	*Gender*	-0.07	0.04	.122	*Gender*	-0.08	0.05	.078
				Synonym judgement	0.02	0.03	.418	*Synonym judgement*	0.01	0.03	.662
				Fluency	-0.01	0.02	.657	*Fluency*	0.002	0.03	.940
				Global Cost	-0.05	0.03	.103	*Global Cost*	-0.04	0.03	.136
				Feature Cost	-0.003	0.02	.915	*Feature Cost*	-0.01	0.03	.603
				Inhibition	-0.01	0.02	.726	*Inhibition*	-0.01	0.03	.653
				Digit span	-0.02	0.02	.485	*Digit span*	-0.02	0.02	.403
								Age Synonym judgement*	-0.02	0.03	.478
								Age Fluency*	-0.03	0.02	.238
								Age Global Cost*	-0.04	0.03	.299
								Age Feature Cost*	-0.03	0.03	.276
								Age Inhibition*	0.01	0.02	.765
								Age Digit span*	0.001	0.02	.947
Model 1 Stats: (F(5,70) = 1.077, p= .381 Total variance explained :7.1%				Model 2 Stats: (*FChange*(6,64) = .866, p= .525 Variance explained relative to model 1: 7.0%				Model 3 Stats: (*FChange*(6,58) = .790, = .582) Variance explained relative to model 2: 6.5%			

Table 8. Hierarchical regression table showing the contributions of the predictor variables to the Mismatch effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

DOI: 10.1177/17470218241309602

model 1				model 2				model 3			
	B	***SE***	***p***		***B***	***SE***	***p***		***B***	***SE***	***p***
intercept	-0.22	0.10	.024	*intercept*	-0.22	0.10	.029	*Intercept*	-0.13	0.13	.327
Age	0.12	0.08	.120	*Age*	0.05	0.10	.623	*Age*	0.01	0.10	.953
Education	0.06	0.07	.422	*Education*	0.05	0.07	.519	*Education*	0.02	0.08	.821
Lifestyle	-0.10	0.08	.180	*Lifestyle*	-0.07	0.08	.372	*Lifestyle*	-0.05	0.09	.599
Social net	0.10	0.07	.183	*Social net*	0.10	0.07	.194	*Social net*	0.12	0.08	.132
Gender	0.12	0.14	.380	*Gender*	0.12	0.15	.401	*Gender*	0.12	0.15	.439
				Synonym judgement	0.05	0.10	.602	*Synonym judgement*	0.03	0.11	.807
				Fluency	-0.05	0.08	.548	*Fluency*	-0.03	0.08	.729
				Global Cost	0.04	0.10	.697	*Global Cost*	0.06	0.10	.548
				Feature Cost	0.02	0.08	.785	*Feature Cost*	0.02	0.09	.807
				Inhibition	0.09	0.08	.248	*Inhibition*	0.10	0.08	.224
				Digit span	-0.04	0.07	.567	*Digit span*	-0.07	0.08	.372
								Age Synonym judgement*	-0.03	0.11	.769
								Age Fluency*	-0.01	0.08	.909
								Age Global Cost*	-0.17	0.11	.136
								Age Feature Cost*	0.003	0.09	.974
								Age Inhibition*	0.01	0.08	.949
								Age Digit span*	-0.05	0.07	.531
Model 1 Stats: (F(5,70) = .967, p = .444) Total variance explained: 6.5%				Model 2 Stats: (*FChange*(6,64) = .553, p = .766) Variance explained relative to model 1: 4.6%				Model 3 Stats: (*FChange*(6,58) = .670, p = .674) Variance explained relative to model 2: 5.8%			

Table 9. Hierarchical regression table showing the contributions of the predictors to the Context effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

DOI: 10.1177/17470218241309602