Do some elderly adults interpret scalar terms like younger adults?

The effect of healthy aging on the processing of the scalar term *some*

Iris van Schaik
s4800265
24 July 2021
Master thesis Taal- en Spraakpathologie (Radboud University)
Primary supervisors: Prof. dr. R. J. Hartsuiker & M. Slim (University of Ghent)
Secondary supervisor: dr. E. Janse (Radboud University)
15622 words (excluding references, tables, and appendices)

Abstract

This thesis studies the processing of the scalar term *some* by healthy old and healthy young adults. The scalar term some has two interpretations: a logical interpretation (some, and possibly all) and a pragmatic interpretation (some, but not all). Many studies have shown that the pragmatic interpretation of the scalar term some is cognitively effortful. The goal of the current study is to replicate these findings and additionally examine whether the healthy old adult group, which was characterized by a decreased working memory ability, would make fewer pragmatic interpretations than the young adult group.

Participants were presented with ‘underinformative’ sentences containing *some* that have differed truth values based on the pragmatic and logical interpretations (e.g., *Some dogs are mammals*). While performing this sentence verification task, participants were also asked to complete a dot task in which they had to recall patterns, which burdened their working memory. The working memory manipulation occurred in three conditions: the no-load condition (participants did not have to recall any patterns), the low-load condition (participants had to recall a simple pattern), and the high-load condition (participants had to recall a complex pattern).

The results showed that the old adult group made significantly more logical interpretations than did the young adult group. Moreover, neither age group showed a clear effect of working memory manipulation on the number of pragmatic or logical interpretations. The results of the working memory manipulation are puzzling in light of the existing theories about the processing of scalar terms. The absence of such an effect indicates that the difference in interpretations between the young adults and old adults cannot be explained by the difference in working memory abilities between these groups. This study proposes that a possible effect of age on general (dual) task performance might be an alternative explanation.

Acknowledgements

I would like to thank all my supervisors, Robert Hartsuiker, Mieke Slim, and Esther Janse, for reading and assessing this master thesis. Robert and Mieke have provided me with many ideas and solutions, and especially Mieke has been my savior when I needed help with the technical aspects. I also want to thank all participants who participated in this study. Finally, I wish to show my gratitude to Olaf Knoester, who has supported me through everything.

Table of contents

1. Introduction

Mr. Bean and his girlfriend, Irma, are sharing a bucket of popcorn in the movie theater. After the movie starts, Irma asks her boyfriend for a bit of popcorn. Mr. Bean hands the bucket over, but it feels strikingly light. She asks Mr. Bean: ‘how much did you eat already?’ When he answers with ‘I ate some of it’, Irma assumes there is still popcorn left for her and that he did not eat all of it yet. Irma’s assumption that not all of the popcorn was eaten by Mr. Bean because he said he ate *some* of it, is a scalar inference or scalar implicature.

These scalar inferences, and specifically how they are processed, will be studied in this master thesis. Interestingly, the assumptions that are made to derive a scalar implicature have been shown to be cognitively effortful (e.g., Bott & Noveck, 2004; Huang & Snedeker, 2009; Van Tiel & Schaeken, 2017). Additionally, there are studies showing that a burdened working memory capacity influences the derivation of scalar implicatures (e.g., De Neys & Schaeken, 2007; Marty & Chemla, 2013; Cho, 2020). In this thesis I will focus on the derivation of scalar implicatures by a certain age-group, in which aging influences cognition: elderly persons.

The rest of this introduction will cover the existing literature on scalar implicatures in general, the processing of scalar implicatures, the effortfulness of scalar implicatures and, finally, cognitive aging.

1.1 What are scalar implicatures?

It is safe to say that we make implicatures, like Irma did, every day. Research about scalar implicatures first appeared around five decennia ago (e.g., Horn, 1972). The main idea is that a speaker does not express more than is necessary and relevant, but is as informative as possible, as described by Grice’s conversational maxims (1975, 1989). So, in the case of (1) below, the listener interprets that John ate some, but not all the pieces of the pie. The fact that John could have eaten all of the pie should be cancelled by the expectation that the speaker would have said all the pieces, and not some of them, if that was the case. This phenomenon arises specifically in the case of scalar terms because, as the name says, these terms exist on a scale of informativeness (Horn, 2004). Using the term *all* is more informative than using the term *some*, which causes the term to have a lower bound meaning (*some and possibly all*) and an upper bound meaning (*some but not all*) (Horn, 2004). These two meanings or interpretations differ based on whether or not the implicature is made and are often called the logical interpretation and the pragmatic interpretation. To demonstrate, the logical interpretation of *some* in (1) is *some and possibly all pieces of the pie,* whereas the pragmatic interpretation of *some* in (1) is *some but not all pieces of the pie*.

(1) John ate some pieces of the pie.

The two interpretations can in certain sentences contrast with each other. For example, in (2) the logical interpretation *some and possibly all dogs are mammals* is correct, whereas the pragmatic interpretation *some but not all dogs are mammals* is incorrect. Sentences such as (2) are often called 'underinformative sentences'.

(2) Some dogs are mammals.

In the case of Mr. Bean and Irma's somewhat comical narrative from the introduction, it is not unlikely that Mr. Bean did in fact eat all of the popcorn, while saying that he ate some of it. This would be correct when he ate *some and possibly all* of the popcorn (the logical interpretation). In that context, Mr. Bean's response would also be underinformative.

1.2 Theories on the processing of scalar implicatures

There are different theories about how scalar implicatures are derived and about how they are processed (Chemla & Singh, 2014). First of all, there are pragmatic and grammatical approaches to the emergence of scalar implicatures. The pragmatic approach assumes that scalar implicatures are a result of pragmatic reasoning (mostly established by Grice, 1975). The grammatical account, on the other hand, claims that scalar implicatures are derived through grammatical operations (Chierchia, 2004; Chierchia, Fox & Spector, 2012).

However, more importantly in the context of this thesis, there are different theories and predictions about how scalar implicatures are processed. The two main theories are the default and the non-default approaches to scalar inferences. As explained above, an implicature or inference has to take place in order to get to a pragmatic interpretation of a sentence with a scalar term. According to the default approach, scalar inferences are made automatically and are cancelled when the context requires it (e.g., Levinson, 2000; Horn, 2004). Non-default accounts, on the other hand, assume that scalar inferences are not made automatically. An example is Relevance theory, which is a context-driven view (Sperber & Wilson, 1986/1995, 2002; Carston, 1998). This approach states that listeners make the scalar inference if it is relevant enough, based on the context. Furthermore, there are some scholars who claim that, even though the fact that the derivation of scalar implicatures overall is quite automatic, this process consists of two different 'steps' (Huang & Snedeker, 2009, 2011; Tomlinson, Bailey & Bott, 2013). The first step is the semantic interpretation (i.e., the logical interpretation), and the second step is the pragmatic interpretation, and the latter is claimed to be cognitively costly.

In both default and non-default theories, the cancellation of either the pragmatic interpretation in a sentence that is only correct with a logical interpretation (default accounts) and the cancellation of the logical interpretation in a sentence that is only correct with a pragmatic interpretation (non-default accounts) are effortful and thus cognitively costly (Bott & Noveck, 2004; Cho, 2020). Yet, some studies report findings that do not fit with either prediction about the effortfulness of the implicatures. An alternative approach, called the constraint-based approach, was introduced by Degen and Tanenhaus (2011, 2014). They carried out three experiments using the ‘gumball paradigm’, where participants saw a gumball machine that had dropped zero to thirteen of the total thirteen gumballs that were inside. Participants then heard the sentence ’You got x gumballs’, where ‘x’ could be ‘some’, ‘some of the’, ‘all of the’, or ‘none of the’. Participants were asked to judge the naturalness of these sentences given the number of gumballs that were dropped on a seven-point Likert scale. They found that *some* was less natural in reference to small sets (1, 2, or 3 gumballs) and in reference to the whole set (13 gumballs). They also found that adding items with exact number words on the ‘x’ in the experimental design lowered the naturalness of items with *some* for small sets (1, 2, or 3 gumballs) but not for intermediate sets (6, 7, or 8 gumballs). The results of their third experiment, which measured reaction times, supported these prior results. The authors argue that their results do not fit either default accounts or non-default accounts and propose that inferences are always derived with the same mechanisms but may appear context-based (effortful) or default (automatic). In other words, the cognitive effortfulness of inferences depends on how many cues for the inference exist and how strong they are. That is, cues like context and/or prosody can facilitate the making of the inferences, though this also depends on their strength.

Another example of a study supporting the constraint-based account is a Magnetoencephalography (MEG) study by Politzer-Ahles and Gwilliams (2015). In this study, short stories were read aloud to participants. These stories were differently enriched with cues for the inference of *some*, and participants answered comprehension questions throughout the procedure. The MEG-results showed that more cognitive effort was required in contexts that provided less support for this inference. In other words, different contexts in which inferences were made called for different amounts of cognitive effort. According to Politzer-Ahles and Gwilliams (2015), these results could hypothetically lend support for non-default theories, since these theories do not presume that all contexts need the same amount of cognitive effort. However, the authors argue that the results fit better with the constraint-based account. Another study that supports this account is by Politzer-Ahles and Fiorentino (2013). They did a self-paced reading task with underinformative sentences and found no significant differences in reading times between semantic and pragmatic interpretations. These authors also interpret their findings as supportive for the constraint-based account, since they found that inferencing is both context-sensitive and potentially effortless.

1.3 Cognitive effort and scalar implicatures

So, is it cognitively effortful for Irma to assume that Mr. Bean did not eat all of the popcorn when he says he ate some of it? Many studies have carried out experiments to test the predictions for processing costs of scalar implicatures by the above-described theories. Most of this is done for the scalar term *some*.

Multiple studies directly tested the non-default approaches of scalar inferencing, and found results that support these theories (e.g., Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny, Katsos & Williams, 2006; Huang & Snedeker, 2009, 2011; Bott, Bailey & Grodner, 2012; Tomlinson et al., 2013; Van Tiel & Schaeken, 2017). Noveck and Posada (2003) looked at the event-related potential (ERP) and reaction time responses to underinformative sentences (e.g., *Some turtles have shells*). They tested French-speaking young adults. They found that participants who responded with false, and thus having made the pragmatic inference (i.e., not *some*, but *all* turtles have shells), had significantly higher reaction times than participants who responded with true (which means having the logical interpretation). The ERP findings supported the fact that inferences in underinformative sentences involve a late and effortful processing. A detailed study by Bott and Noveck (2004) replicated the reaction time results of Noveck and Posada (2003). They employed a truth-value-judgment task in four different experiments, where French speaking participants had to judge underinformative sentences (e.g., *Some elephants are mammals*), and where the answer given (true of false) was indicative for the type of interpretation they had (logical or pragmatic). In the first three experiments, the authors examined if counterbalancing the answer-interpretations combinations by using adjusted stimuli (e.g., *Mary says the following is true/false: some elephants are mammals*) and giving explicit instructions to the participants to have a certain interpretation significantly changed the results. These factors did not, in fact, change the results, which indicated that reaction times of the 'false' responses (i.e., the pragmatic interpretation) were significantly longer than the 'true' responses (i.e., the logical interpretation). In their fourth experiment, the authors specifically tested whether the pragmatic interpretation was cognitively effortful. Two conditions were added to their prior design in order to manipulate the cognitive resources; in the Long Condition, participants had 3 seconds to respond and in the Short Condition, participants had 900 ms to respond. The authors speculated that a shorter time to react would limit the cognitive resources available to make the pragmatic inference. Results showed that participants were indeed less likely to have a pragmatic interpretation in the Short Condition, supporting the claim that the inference of the scalar term *some* is cognitively effortful.

A reading time study by Breheny et al. (2006) contradicts the default accounts as well, and specifically supports the context-driven approach. They tested Greek-speaking university students and found longer reading times for small stories in conditions where the scalar implicature was inevitable because of the context. Further evidence comes from an eye-movement study done by Huang and Snedeker (2009), in which a visual-world eye-tracking paradigm was used. The goal of using eye-movements to study scalar inference was to indirectly and temporally tap into comprehension. Another advantage was that the measured movements were unconscious. In the three experiments in this study, participants heard stories in which four persons, two boys and two girls, possess two types of items in different quantities. This was also visually displayed. The participants, who were monolingual English speakers and university students, then heard a command to point at one of the four pictures (e.g., *Point to the girl that has some of the socks*), during which their eye movements were recorded. The experiments showed that processing of the sentences was quicker when participants assigned the logical interpretation than when they assigned the pragmatic interpretation, suggesting that the pragmatic interpretation involves processing costs. Together with the fact that participants actually made the pragmatic inference when that was not necessary based on the context, the authors conclude that pragmatic inference is not the default, but that their results do not fit any context-driven accounts. They propose that the results fit better with a model of language processing that has distinct levels, where the semantic level comes after the phonological level and before the pragmatic level. Semantic representations, thus, must precede the pragmatic interpretation.

A somewhat different body of research confirming the non-default approach, which proposes that processing effort is necessary for making a scalar inference, are studies where cognitive resources are manipulated during scalar inferencing. Often this is done with a dual task, where participants have to perform a task that calls upon working memory while concurrently doing the scalar task. In these studies, it is said that the working memory task takes up cognitive resources that are also needed for the inference of the scalar term. More complex working memory tasks result in less working memory capacity available for inferencing. Most of these studies show that when working memory ability is burdened or limited, fewer pragmatic interpretations are made, supporting the claim that cognitive effort is needed for inferencing (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard & Schaeken, 2011; Marty, Chemla & Spector, 2013; Marty & Chemla, 2013; Van Tiel, Pankratz & Sun, 2019; Van Tiel, Pankratz, Marty & Sun, 2019; Cho, 2020). There are several ways in which cognitive effort was manipulated in these studies. Bott and Noveck (2004), for example, shortened the time to respond to the underinformative sentences in their study. However, most studies decrease the working memory capacity that participants have available for the scalar inferencing by employing a dual task. This dual task often consists of the sentence verification task where participants judge the underinformative sentence and a concurrent task that demands working memory resources, for example a dot task. This is a task where participants remember a pattern of a certain number of dots.

The first to examine the role of working memory in scalar inferencing like this were De Neys and Schaeken (2007). They carried out a sentence verification task with underinformative sentences (e.g., *Some eels are fish*) and implemented a simultaneous dot task to manipulate cognitive resources. Participants judged the underinformative sentences while having to remember patterns of dots, which were presented in 3 by 3 grids. Working memory load was manipulated differently, by presenting the sentences in a load condition and a control condition. During the load condition patterns consisted of four dots in a complex pattern and during the control condition, patterns consisted of three dots in a simpler pattern (exclusively on a horizonal or vertical line in the grid). They found that participants, who were Dutch-speaking university students, made 5.7% more pragmatic interpretations in the control condition than in the load condition (this difference was statistically significant). So, they found that fewer scalar inferences were made when cognitive resources were minimally spent on the dot task, than when they were highly spent on the dot task. This indicates that scalar inferences are cognitively effortful and are not automatic, supporting non-default accounts. A study by Marty and Chemla (2013) replicated the results found by De Neys and Schaeken (2007) by applying the same experimental design. They tested young adults which were native speakers of French. Additionally, they tested sentences containing *only some* (e.g., *Only some politicians are corrupt*), which did not show the same effect of working memory manipulation as the sentences with exclusively *some* did. Marty et al. (2013) demonstrated similar results with sentences containing bare numerals, which actually behaved the opposite of sentences with *some* (so, more inferences in conditions that burdened working memory more). Moreover, studies done by Van Tiel, Pankratz and Sun (2019), Van Tiel, Pankratz, Marty and Sun (2019), and Van Tiel, Van Miltenburg, Zevakhina and Geurts (2014) show that some other scalar terms (e.g., *might*, *or*, *scarce*, etc.) do not yield the same effect of working memory manipulation as does *some*.

However, other studies looking at the processing of scalar terms show very different results (e.g., Feeney, Scrafton & Duckworth, 2004; Grodner, Klein, Carbary & Tanenhaus, 2010; Holtgraves & Kraus, 2018). Feeney, Scrafton and Duckworth (2004) dispute the study by Noveck and Posada (2003), in which reaction times turned out to be longer for participants who responded pragmatically to underinformative sentences than for participants who responded logically to these sentences. They argue that the conclusion made by Noveck and Posada (2003), which was that the results support the non-default Relevance theory, is not justified, because it is unknown what processes were involved in these reaction times. The authors reason that the possibility of response strategies of the participants, both for logical and pragmatic responders, might have caused the differences between the groups. In order to avoid participants having response strategies, Feeney, Scrafton and Duckworth (2004) set up a similar experiment with fewer trials which contained more different types of trials. They tested English-speaking university students and measured the reaction times of the *true* or *false* responses to the trials (that consisted of underinformative sentences and fillers). The authors compared reaction times of *true* responses with *false* responses to sentences with *some* that were underinformative, of participants who did not seem to show a certain response strategy (i.e., they only analyzed participants who gave logical as well as pragmatic answers and excluded participants from this analysis who answered either always pragmatic or always logical). These results did not show a significant difference in reaction time between the logical and pragmatic response, disproving the results of Noveck and Posada (2003). Additionally, they found that the reaction time of logical responses to underinformative sentences with *some* was significantly longer than the logical response to sentences with *some* without a possible pragmatic interpretation. The authors interpret this as the participants having to inhibit the pragmatic interpretation in the underinformative sentences. Thus, they claim they demonstrated processing costs for the logical interpretation instead of the pragmatic interpretation, supporting the default theories.

Another study with divergent results is by Grodner et al. (2010), who criticized the design used by Huang and Snedeker (2009). They discuss multiple aspects of the design and how these could have caused Huang and Snedeker's (2009) findings that processing of sentences with a pragmatic reading of *some* took significantly longer than the processing of other sentences. For example, Huang & Snedeker (2009) used more sentences containing exact numbers as stimuli than sentences with *some of*, which might have resulted in less acceptability of sentences with *some of* (Degen, Reeder, Carbary & Tanenhaus, 2009; Degen & Tanenhaus, 2014). Consequently, Grodner et al. (2010) adjusted Huang and Snedeker's (2009) experimental design in order to circumvent the shortcomings that they found. They looked at the eye movements of 25 English speaking adults that heard (underinformative) sentences with *some* and were presented with a display of six possible referents. The results demonstrated no evidence of a delay for sentences where the inference of *some* was made and are thus inconsistent with the results from Huang and Snedeker (2009). The authors claim that context plays a big role in their results, which can be seen as supportive of the constraint-based account by Degen and Tanenhaus (2014).

Lastly, Holtgraves and Kraus (2018) investigated the processing of five different scalar terms (*some*, *sometimes*, *like*, *good*, and *possible*) in different conversational contexts with ERP during a self-paced reading task. The authors looked at conversational contexts that differ in politeness, since some studies have shown that by hearing sentences with *some* in a potential 'face-threatening' context, e.g., *Some people hated your speech*, participants were more likely to accept the logical interpretation (that some and possibly all people hated the speech) because *some* might be used by the speaker in a polite way (Bonnefon, Feeney & Villejoubert, 2009; Bonnefon, De Neys & Feeney, 2011). The awareness of the listener that *some* is used politely, blocks the pragmatic meaning (which would be that some, but not all people hated the speech). Additionally, it seemed that in the face-threatening conditions, the reaction times of sentences that were judged with a logical response were longer than those of sentences judged with a pragmatic response. This would mean that in that context, the logical interpretation was cognitively effortful, and not the pragmatic interpretation. Holtgraves and Kraus (2018) replicated these findings with ERP. They tested forty students and presented them with scenarios in different face-threatening contexts. Participants went through these stories in a self-paced manner, and then judged a statement based on the scenario. The results demonstrated a bigger ERP-effect (P300) in sentences with logical interpretations than in sentences with pragmatic interpretations. This difference turned out to be bigger in the face-threatening contexts. The authors conclude from this that the pragmatic interpretation is expected and not the logical interpretation, which is predicted by the default accounts, but that this can be dependent on the context.

In short, the literature offers several accounts about the processing of scalar inferences. A considerable number of studies show that the pragmatic interpretation of the scalar term *some* is cognitively effortful (non-default approaches). Some studies dispute these findings, either showing that in fact the logical interpretation is cognitively effortful (default approaches) or that scalar terms can both appear effortful or automatic depending on the context (constraint-based approach). The objective of this thesis is to study scalar inferencing in a specific age group: healthy old adults. So, the next section will be a short overview of the general literature on cognitive aging and existing research on pragmatic reasoning by old adults.

1.4 Working memory and healthy cognitive aging

Cognition changes with age (Salthouse, 2001, 2004; Harada, Love & Triebel, 2013). Examples of general cognitive domains that decline with age are processing speed, attention, memory, visuospatial abilities/construction, and executive functioning (Albinet, Boucard, Bouquet & Audiffren, 2012; Harada et al., 2013). During life, grey and white matter volume decline, white matter changes, and neurotransmitter levels lower, which all might contribute to these cognitive changes (Harada et al., 2013).

Working memory, which is generally referred to as the ability to retain information while at the same time being able to manipulate that information, also becomes less effective as healthy people become older (e.g., Salthouse, Mitchell, Skovronek & Babcock, 1989; Salthouse, 1990; Head, Rax, Gunning-Dixon, Williamson & Acker, 2002; Fisk & Sharp, 2004; Reuter-Lorenz & Sylvester, 2005; Lubitz, Niedeggen & Feser, 2017; Klencklen, Lavenex, Bradner & Lavenex, 2017; Jarjat, Portrat & Hot, 2019). Working memory is seen as the intermediate state between sensory memory, in which perceptual information remains active for a short amount of time, and long-term memory, which is the long-term storage of information that is not active (Meeter & Hendriks, 2020). In Baddeley's (2003) model, working memory consists of one central executive component and three buffers for limited storage of specific information: the visuo-spatial sketchpad, the phonological loop, and the episodic buffer. The central executive interacts with all three buffers and regulates the executing and coordinating of the information that is being altered at that moment. The buffers, then, act as temporal storage and are modality specific. The visuo-spatial sketchpad temporally holds visual information, and the phonological loop temporally holds verbal information. The episodic buffer combines spatial, visual, and verbal information with chronological ordering into integrated episodic units. The episodic buffer can temporally save information of the visuo-spatial sketchpad and the phonological loop when the space in those buffers is taken up. Several working memory tasks show an effect of aging, such as letter rotation and types of span tasks (for example reading span, computation span, and line span), and the amount of information that can be stored (the 'span' in the span tasks) is among the aspects that are affected most by aging (Brickman & Stern, 2009).

According to Harada et al. (2013), language ability is generally unaffected by aging, with some exceptions (e.g., visual confrontation naming, naming of a common object, and verbal fluency). However, as working memory is linked to (pragmatic) reasoning (Kyllonen & Cristal, 1990; Kemtes & Kemper, 1999), one might expect that pragmatic reasoning is affected in older adults. This would mean that, if Irma were to be an old lady with working memory problems, she possibly might not have been surprised if Mr. Bean handed her an empty bucket. Irma might, in this case, have logically interpreted Mr. Bean's response that he ate some of the popcorn as 'some and possibly all of it', since she would not have made the scalar inference.

Until now, there are no studies investigating scalar inferences by healthy older adults with a decreased working memory capacity. There are studies that have found an effect of aging on verbal and/or inferential reasoning, which is often linked to a decline in working memory abilities (e.g., Cohen, 1981; Salthouse, Mitchell, Skovronek & Babcock, 1989; Hamm & Hasher, 1992; Bielak, Hultsch, Kadlex & Strauss, 2007). Verschueren, Schaeken and Verbrugge (2006) did examine pragmatic counterexamples in older adults. These counterexamples are examples based on one's background knowledge that can cancel a conditional statement. The authors studied disabling conditions and alternative causes as counterexamples. For example, the conditional statement 'if you water a plant, it will stay green' can be cancelled by the answer to the question 'a plant is watered, will it stay green?'. A disabling condition based on background knowledge can cause a negative response to this question. The authors found that older adults did not differ significantly from younger adults for this type of conditional reasoning, while working memory performance was significantly lower for older adults. Also, a study by Verschueren, Schaeken and d'Ydewalle (2004), which carried out a thinking-aloud experiment with and without a working memory burden by a dual-task, confirms that this type of reasoning specifically involves working memory. The authors explain their results by the fact that older adults are more acquainted with conversational implicatures, which compensates for the decreased working memory capacity. This concept, that certain skills and information are built up with age and experience (i.e., 'crystallized abilities') and are not lost by aging, has been recognized for aspects like intelligence, knowledge, procedural skills, grammar, and vocabulary (Brown et al., 2011; Gazes et al., 2020). Indeed, Verschueren et al. (2006) interpret their findings as 'crystallized pragmatics', that compensate for the working memory deficits in the old adults. Another study in which old adults seemed to rely more on built up crystallized abilities than young adults is by Lopukhina, Laurinavichyute and Malyutina (2020). They tested syntactically ambiguous sentences that were unambiguous in terms of their contents (e.g., *Rimma dressed the child of the writer, who published a popular novel*). They found that young adults made more mistakes than did the older adults, indicating that the older adults depended more on the semantic information, whereas the younger adults depended more on the structural information.

1.5 Current study

1.5.1 Research question and objectives

Many studies found that, by manipulating cognitive load, a higher working memory load burdens the pragmatic interpretation of scalar terms. Since working memory seems to be less effective in old people compared to young people, it could be expected that old people make fewer scalar inferences than young people (based on the studies that burdened working memory load by manipulation). This effect was not found for conditional reasoning by Verschueren et al. (2006), but has, to date, not been researched for scalar terms. Apart from the fact that this has not been explored, it is relevant to do so because most studies researching the effortfulness of *some* test young adults (often students). These studies then draw their conclusions based on that young sample, whereas the findings might not be generalizable to older age groups. Thus, in this thesis, the goal will be to compare scalar inferencing between healthy young and healthy old adults. The research question of my thesis will be as follows:

To what extent do healthy old adults differ from healthy young adults in their interpretations of the scalar term *some*, and what is the effect of a working memory load on this difference?

The main goal of this study is to gain insight into how working memory can play a role in pragmatic inferencing and how this changes with age. A complementary objective is to replicate the study done by De Neys and Schaeken (2007), who concluded for healthy young adults that scalar inferences involve effortful processing and therefore called into question the prediction from the default theories for scalar inferencing.

1.5.2 Hypotheses

Following De Neys and Schaeken (2007) and others who found that the pragmatic interpretation of the scalar term *some* is effortful, one would expect that the more working memory is burdened, the less scalar inferences will be made. Consequently, it might be expected that old adults will make fewer scalar inferences than young adults if the old adults show a significantly decreased working memory ability compared to the young adults. However, according to default approaches, one would expect the opposite, namely that the logical interpretation of *some* is effortful. In that case, more scalar inferences will be made in conditions in which working memory is burdened more, and old adults (if they have a significantly worse working memory ability) will make more pragmatic interpretations than young adults. The constraint-based account will not be able to be tested in this study, as the scalar inference will not be studied in different contexts.

Another way to test the above mentioned theories is to look at the relationship between working memory burdening and the type of interpretations that participants have (pragmatic or logical). It may be the case, as is described in the previous paragraph, that a more burdened working memory leads to either more logical or more pragmatic interpretations. However, participants may also engage in a so-called ‘trade-off’. That is, participants may perform worse on the working memory task (that is used for burdening working memory) when having a certain interpretation that is effortful (based on the theories either logical or pragmatic). So, if the pragmatic interpretation is effortful (non-default accounts) one would expect a strong relation between the number of pragmatic answers and the number of mistakes on the working memory task. If the logical interpretation is effortful (default accounts) one would expect a strong relation between the number of logical answers and the number of mistakes on the working memory task.

Based on the study by Verschueren et al. (2006), another outcome could be that the scalar term *some* might be ‘crystallized’ in the old adult participants. This would mean that old adults would not perform very differently from young adults (neither fewer pragmatic inferences nor fewer logical inferences, as predicted by the default and non-default accounts), as the latter group compensates for their working memory deficits with their crystallized skills and/or knowledge that they built up over the years. The working memory manipulation would, in this case, have a bigger effect on the young adults than on the old adults, as the old adults are less dependent on their working memory capacities than the young adults.

1.5.3. Study design

In order to answer the research question and test the hypotheses posed above, two groups of participants were tested on their processing of the scalar term *some*. One group consisted of healthy young adults who are native speakers of English and the other group consisted of healthy old adults who are also native speakers of English. Participants judged underinformative sentences in three conditions where the working memory load was differently manipulated: a no-load condition, a low-load condition, and a high-load condition. This came down to a 2 by 3 design, with Age Group as a between-subject variable and Working Memory Manipulation as a within-subject variable. The dependent variable was the number of times participants gave the ‘true’ answer, which represents the logical interpretation of the scalar term *some*. This way, information would be acquired about how often and when participants did not make a pragmatic inference while processing the underinformative sentences, and how often and when they did (since it is a binary choice).

The current study employed the same low-load and high-load conditions as De Neys and Schaeken (2007), plus an extra condition: the no-load condition. This condition, in which participants did not have to recall any patterns, was added because the goal of the current study was to test old adults, and it seemed interesting to compare the performance of young and old adults on the sentence verification task without any working memory burdening.

In order to demonstrate that the old adult group and young adult group differed from each other with regard to working memory, participants performed a spatial storage task. This task, the Corsi block-tapping task, specifically measures the type of working memory that was manipulated in the dot task. A survey was administered to record participant information, allowing in- and exclusion based on age, hearing/vision problems, and (a history of) medical problems. Additionally, all old participants were screened for cognitive problems, since the main goal was to study healthy aging. The survey was also used to document the control variables gender, education, and country of origin. Further, participants were asked to answer some questions about the experimental materials, so as to potentially exclude potentially deviating items.

Finally, in this study, only the scalar term *some* will be studied since this scalar term has been researched most. In addition to this, a narrow focus on *some* is justified by the fact that not all scalar terms and their processing are comparable with each other (Van Tiel, Miltenburg, Zevakhina & Geurts, 2014; Van Tiel, Pankratz & Sun, 2019; Van Tiel, Pankratz, Marty & Sun, 2019). This would mean that by using different scalar terms, within-study results would be difficult to compare. Relatedly, Janssens and Schaeken (2016) did not find any processing costs for the scalar terms *but*, *so*, and *nevertheless*. Moreover, 'old adults' will be defined as persons of 65 years or older in this study (United Nations, Department of Economic and Social Affairs, Population Division, 2019).

2. Method

2.1 Ethics statement

This study has been approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University. All participants gave their informed consent to participate in this study.

2.2 Participants

A total of 142 participants participated in this study, which was conducted online. Based on the scores on the cognitive screening and medical information that participants provided in the survey, 16 participants were excluded from the study, bringing the total number of participants to 126. All participants were recruited via Prolific (www.prolific.co) and received a compensation of approximately €10,- per hour for their time. Prolific offers custom prescreening, with which participants with certain characteristics can be excluded from the experiment. For the group of old adults, age was set to minimum 65 years and maximum 85 years, for the young adults this was set to minimum 20 years and maximum 35 years. For participants in both groups, first language was set to English, participants had to have no language-related disorders (i.e., no reading difficulty, writing difficulty, or any other language-related disorder), and participants had to have normal or corrected-to-normal vision (i.e., they had to be able to see color normally, and if they needed glasses, they would have to be wearing them or contact lenses).

See Table 1 below for descriptive information for each age group, and Figure 1 below for the residence of the participants at the time of the study. Participants in the group ‘other’ resided in Australia, Israel, and different countries in Europe.

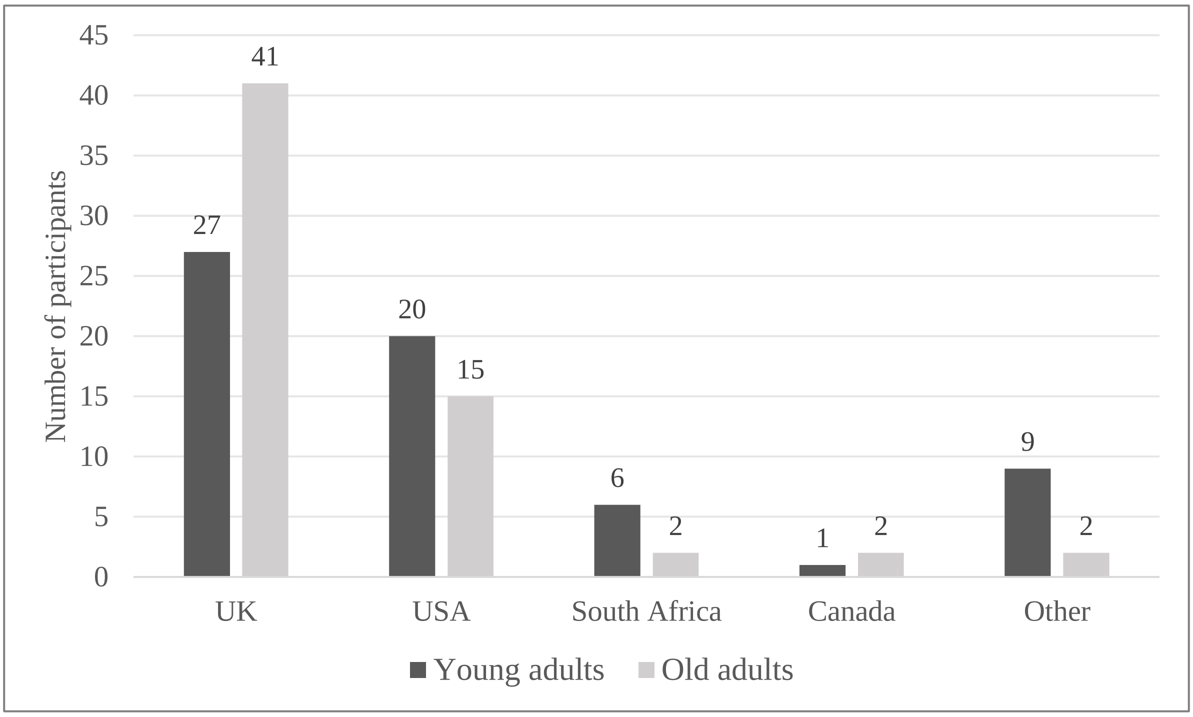
Table 1

Participant Characteristics

	Young adults (*N* = 63)		Old adults (*N* = 63)	
Only speaks English	*N* = 33		*N* = 38	
Gender	29 f; 34 m		28 f; 35 m	
	Mean (*SD*)	Min - Max	Mean (*SD*)	Min - Max
Age (in years)	28.25 (4.45)	20.00-35.00	68.84 (3.17)	65.00-77.00
Formal education (in years)	15.51 (4.85)	0.00-27.00	15.92 (3.34)	8.00-25.00

Figure 1

Residence of the Participants at the Time of the Study (Per Age Group)



2.3 Materials

2.3.1 Sentence verification task

The sentence verification task used in this experiment was based on the task used in De Neys and Schaeken (2007). Participants were asked to judge underinformative sentences with true or false. These underinformative sentences had a general structure, consisting of a category and an example of that category (e.g., *Some bees* (example) *are insects* (category)). Underinformative sentences have two possible interpretations, and the way participants judge these sentences reveals the interpretation they have (logical or pragmatic). Participants were told to click 'true' if they believed the sentence on the screen to true, and to click 'false' if they believed the sentence to be false. Participants judged ten target sentences in each of the three working memory load conditions (see 2.2.2).

Apart from these target sentences, control sentences were presented that also acted as fillers. The fillers consisted of sentences with *some* that were clearly true (e.g., *Some fish are goldfish*) or clearly false (e.g., *Some beetles are flowers*) and sentences with *all* that were clearly true (e.g., *All parrots are birds*) or clearly false (e.g., *All flowers are pansies*, *All wasps are trees*). There were ten control sentences in each working memory load condition. So, in each condition, participants judged twenty sentences, and across the whole experiment a total of sixty sentences. The same forty sentences from De Neys and Schaeken (2007) were used in order to replicate their study (see their Appendix). However, since the current study employs a third working memory condition that was not present in De Neys and

Schaeken (2007), ten extra target sentences and ten extra fillers were created. See Appendix 1 for all sentences used in the current study.

Similar to De Neys and Schaeken (2007), target sentences and fillers were presented in a random order for each participant and per condition. Apart from this, all three sentence sets were used an equal number of times in all three working memory conditions. The sentence verification task, as well as the dot task, online cognitive screening, survey, and feedback and comments page, was programmed in PennController for Internet Based Experiments (PCIbex; Zehr & Schwarz, 2018). This project offers an interface along with JavaScript-based syntax to create online experiments. Experiments can also be run from their “farm”.

2.3.2 Dot task

In order to introduce a working memory load manipulation, participants had to recall a previously shown dot pattern after judging the sentences. Materials were again based on those used in De Neys and Schaeken (2007). Dot patterns were presented for 850 ms in 3 by 3 grids. The variation of working memory manipulation was based on the complexity of the to be remembered patterns. In the no-load condition, participants were shown an empty 3 by 3 grid; they did not have to recall any pattern. In the low-load condition, participants had to recall a simple pattern consisting of three dots. These three dots were presented horizontally or vertically, in a row. The complex patterns in the high-load condition consisted of four dots which were scattered over the 3 by 3 grid. Patterns within conditions were presented in a fixed order, in order to stay as close to De Neys and Schaeken’s (2007) design as possible. See Appendix 2 for the exact patterns and order within conditions.

For the dot task, participants were instructed to recall the patterns as correctly as possible. One difference from De Neys and Schaeken’s (2007) design was that participants did not receive feedback on whether they reproduced the pattern correctly. The patterns that participants submitted were documented, enabling an analysis of which patterns were recalled incorrectly.

The dot task, as used here, is a spatial memory task (e.g., Ichikawa, 1983; Bethell-Fox & Shepard, 1988; Miyake, Friedman, Rettinger, Shah & Hegarty, 2001; De Neys & Schaeken, 2007). This task was chosen by De Neys and Schaeken (2007) in order to load working memory, because it particularly burdens the executive component of working memory. De Neys and Schaeken (2007) originally only employed two working memory conditions; one where participants had to remember a simple pattern of three dots and one where participants had to remember a complex pattern of four dots. A study by Klencklen et al. (2017) showed that the amount of memory load had a significant effect on age-related changes in working memory performance. Hence, this is why the no-load condition, where participants do not have to recall any pattern, was added in this study. This way, performances in the conditions with a working memory load could be compared to a baseline, and any possible floor effects of the old adult participants might be avoided.

2.3.3 Corsi block-tapping task

The Corsi block-tapping task (Corsi, 1972; Berch, Krikorian & Huha, 1998; Kessels, Van Zandvoort, Postman, Kappelle & De Haan, 2000) was used to measure the working memory capacity of the participants. The goal of this task was to determine differences between the old adult group and young adult group. In the Corsi block-tapping task, a sequence of highlighted blocks is presented to participants, who have to recall this sequence by selecting the same blocks. The sequences get incrementally longer, and the task stops when the participants cannot recall the sequence correctly anymore. The highest number of blocks that can be remembered is called the Corsi span. The Corsi task measures spatial working memory (Kessels, Van Den Berg, Ruis & Brands, 2008; Power, 2017), and multiple studies have found age-related differences on this task (e.g., Brown, 2016; D'Antuono, Maini, Marin, Boccia & Piccardi, 2020).

The design and procedure of the online Corsi task in the current study was based on Brunetti, Del Gatto and Delogu (2014), who implemented the Corsi block-tapping task for tablets under the name of 'eCorsi'. The implementation of the task was done via PsyToolkit (Stoet, 2010, 2017), an online software package for creating and running online experiments, which offers a general design structure of the Corsi block-tapping task.

2.3.4 Online MoCa test

In order to exclude older participants with cognitive problems from this study, a cognitive screening was needed. Since there were no options available to use an existing online cognitive screening, it was necessary to program one for online use. The Montreal Cognitive assessment (MoCa) test (Nasreddine et al., 2005) was chosen. This test is a concise screening tool for measuring cognitive decline. The screening covers multiple cognitive domains: executive functioning, visual skills, attention, concentration, working pace, language, short term memory, and orientation. The MoCa test is generally used to screen for illnesses characterized by moderate to severe cognitive decline (e.g., Alzheimer's disease, Parkinson's disease, stroke, Multiple Sclerosis, etc.), but can also be used to detect a milder form, namely Mild Cognitive Impairment (MCI).

A practical reason for using the MoCa test was its simplicity compared to other cognitive screenings, allowing a relatively uncomplicated online implementation. Indeed, the website of the MoCa test even offers a blind version of the MoCa, which is usually administered over the phone and thus does not include the visual elements (Wittich, Phillips, Nasreddine & Chertkow, 2010). The visual elements that are absent in this version are the Alternating Trial Making test (drawing a line from alternating points), the two Visuoconstrucional Skills tests (copying a three-dimensional cube and drawing a clock set on a certain time), and the Naming test (naming figures of certain animals). For the online implementation in this study, the visual tests were also removed, except for the Naming test which could be performed online (whereas the rest could not). Moreover, it was not possible to perform two items of the Orientation test (asking the participant for their 'place' and 'city') due to ethical reasons, which subtracts two points from the total score. This brings the total score of the online implementation of the MoCa used in this study to 23. Participants reporting an amount of 12 years or less of formal education received one point on top of their final score (Nasreddine et al., 2005). Based on the scoring of the blind version of the MoCa test and the fact that the test is adjusted for online implementation it can be determined that a score of 18 or more is considered normal. For this reason, all data of participants in the old adult group with a MoCa test score lower than 18 were deleted from the analysis. See Appendix 3 for the distribution of points per task.

Some instructions and items of the MoCa test had to be presented aurally to participants in order to implement the test online. For this, a native speaker of English was recruited via Prolific, who recorded herself saying out loud the components of the MoCa test that required audio. See Appendix 4 for all components that were recorded. The recruited native speaker of English was a 29-year-old female from Victoria (British Columbia), Canada, who moved there from Ontario at the age of 18. This speaker's accent did not appear to be influenced by other acquired languages (she only casually learned French in school from age 8 to 15). All audio files were prepared for online implementation by editing them to the appropriate length and combining certain files to one file, using version 3.0.2 of the free recording and editing software Audacity® (Audacity Team, 2021).

2.3.5 Survey and feedback/comments

Participants were asked for their age, gender, number of years of formal education, current residence, native language, second languages, hearing/reading problems, and hearing/reading problems in the experiment. See Appendix 5 for how the survey was presented to the participants.

Finally, participants were provided with the opportunity to give feedback and comment on the study. Two questions on this page contained a sentence from the sentence verification task (*Some robins are birds* and *some ants are insects*). The participants were asked whether they answered true or false for these sentences, and why. This could give more insight in how participants reasoned. Specifically, for the *some robins are birds* sentence, it would be interesting to see if participants involved their general world knowledge in their judgment (one could argue that not all robins are birds, since there are people named Robin). The third question was about the stimuli used in the sentence verification task. Participants were shown a table of all categories and members of those categories used, and they were asked to write down the words that were not familiar to them during the experiment. In the fourth and last question they were asked if they had any other comments. See Appendix 6 for how the feedback page was presented to the participants.

2.4 Procedure

2.4.1 Young adults

The procedure for the young adult group was the same as for the old adult group, with the only difference being that the old adults had to complete a cognitive screening. On Prolific, participants received the link to the study in PCIbex. The first thing they saw was a question that asked: *Are you a native speaker of English?* Participants who answered yes to this question saw another screen where they had to indicate whether they had ever been clinically diagnosed with certain medical problems. These medical problems were shown on the screen, and included any form of dementia, Parkinson's disease, depression, schizophrenia, stroke, and any form of head trauma. These are all common illnesses that can cause cognitive problems, and which the MoCa test screens for. Having a negative answer to either of these two questions disabled participants to continue with the experiment.

If a participant passed the two questions above, they moved on to a welcome screen with a general explanation of what would be expected from them during the experiment. Then, a link was shown that the participants had to click in order to be redirected to the PsyToolkit environment, where they were presented with a general study information page, an informed consent page, instructions for the Corsi block-tapping task, and the Corsi task itself. After a countdown from three, participants saw a total of 12 purple blocks on their screen. The sequences they had to remember are shown by briefly and consecutively changing the color of several blocks to yellow. Blocks flashed yellow for 500 ms and the time between this flashing was 1000 ms. Afterwards, the participants had to recall the sequence by clicking the blocks on the screen in the correct order. The participant needed to click a green block with the word 'done' written inside of it to submit their answer and received feedback on whether they recalled the sequence correctly (by being shown a smiling or a frowning emoticon). The first sequence started with two blocks. If a sequence was recalled correctly on the first try, the participant directly moved on to the next sequence (which was always a sequence with one more block than the previous one). If a sequence was recalled incorrectly on the first try, the participant got to recall a new sequence with the same number of blocks again. The task ended when the second try was not executed correctly. The number of blocks of the longest sequence that was remembered by the participant is the Corsi block span, which was also displayed on the screen.

After the Corsi block-tapping task, participants were asked to enter their Prolific ID. Subsequently, they were shown another link they had to click in order to continue with the experiment in PCIbex. Here, they were first asked to enter their Prolific ID again. This was needed in order to link Corsi block-tapping results to the results of the rest of the experiment for each participant. Then followed the instructions for the dual task (the sentence verification task and the dot task). In these instructions the conditions were explained, and it was stressed that patterns had to be correctly recalled. Given the length of the study and the burden on participant's energy level, participants were urged to take a break between segments. This was stated in the instructions, along with that it would shown on their screen when they could take a break. The instructions were followed by three test items, one for each working memory condition. These test items reflected the real target and filler sentences in the conditions: a 3 by 3 grid was shown for 850 ms. Some or none of the boxes in this grid were checked, depending on the working memory condition. After the disappearance of the grid, the sentence was presented on the screen with two buttons underneath saying ‘TRUE’ and ‘FALSE’. The sentence disappeared once the participant made their choice, and then an empty 3 by 3 grid appeared where the previously seen pattern could be reproduced. A button with ‘continue’ then brought the participant to the next item. See Figure 2, 3, and 4 below for examples of the patterns that were displayed in the different working memory load conditions. Once the dual task was completed, participants were presented with a survey containing ten questions, and subsequently a feedback and comments page with four questions.

Figure 2

The Empty Grid Shown in the No-Load Condition

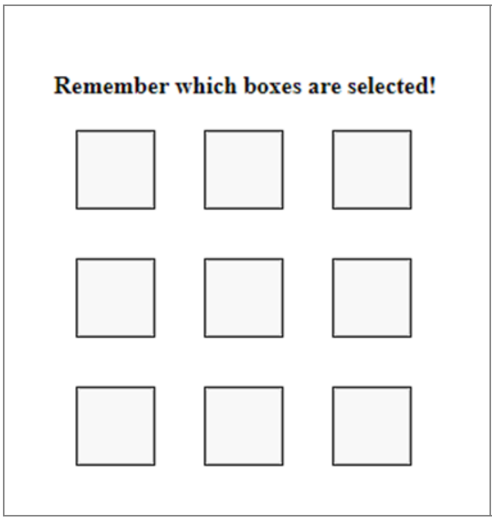


Figure 3

Example of a Dot Pattern Shown in the Low-Load Condition

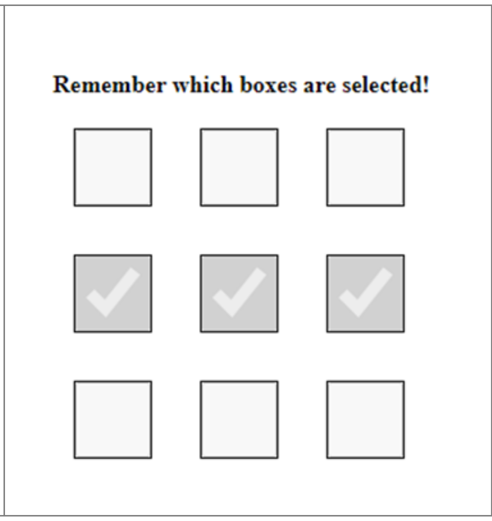
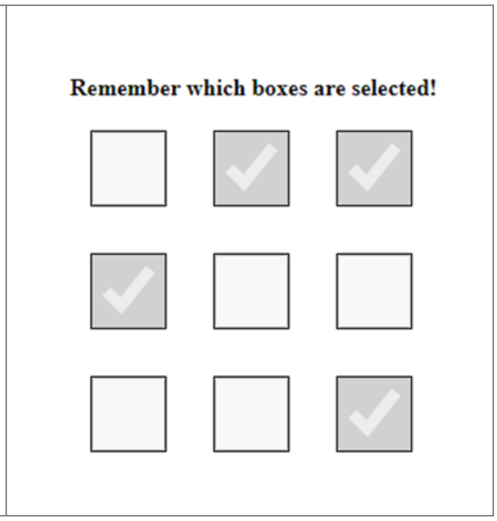


Figure 4

Example of a Dot Pattern Shown in the High-Load Condition



2.4.2 Old adults

The procedure for the old adult group was the same as for the young adult group (as described above), but they additionally completed a cognitive screening to ensure that they did not have cognitive problems. The old adult participants also had to test their sound volume and microphone at the start of the experiment (after the question about English as their native language), as this would be important for the cognitive screening. The online implementation of the MoCa test was programmed after the completion of the dual task and before the survey page, and the order of the items was copied from the normal version of the MoCa (see Appendix 3). All instructions were primarily based on the normal version and the phone version of the MoCa, with some adjustments due to the online design of the tasks.

The cognitive screening started with a general introduction and explanation of the screening. Microphone and sound volume were again tested here, so participants would be able to hear and produce items well enough. Then the browser of participants went into full screen mode. This was necessary for certain questions in the screening for which the use of external resources (like a calculator, calendar, access to the internet, etc.) was disallowed. The first part of the screening was a naming task, where participants were shown three pictures of animals (a lion, rhinoceros, and camel). They were asked to type the name of each animal under each picture. Then followed the memory part of the screening, that involved hearing five words (‘face’, ‘red’, ‘daisy’, ‘church’, and ‘velvet’) and repeating these. The audio recording of the five words was played to the participants, after which they had to record themselves saying as many of the words as they could remember (order did not matter). This process was then repeated again, and it was stressed that the words had to be recalled at the end of the screening (for the delayed recall task).

The next part of the screening consisted of a forward digit span task and a backward digit span task. Two sequences of numbers, ‘21854’ and ‘742’, were played aurally to the participants. The first sequence had to be repeated in the same order it was heard, but the second sequence had to be repeated backwards (so the answer is ‘247’). Participants had to record their answers. Next was another attention task, where a list of letters was played aurally at approximately one letter per second. The objective for the participant was to press the space bar every time they heard the letter ‘A’, and do nothing if they heard any other letter. The last attention task involved serial subtraction of seven starting at 100. It was stressed that using a calculator was not allowed, and that participants should continue subtracting seven if they felt like they had made a mistake. Participants typed their answers.

The next task involved repeating two sentences: ‘I only know that John is the one to help today’, and ‘The cat always hid under the couch when dogs were in the room’. These sentences were played aurally, and the participants had to record themselves repeating the sentences exactly. The other language task was naming as many words as possible that start with the letter ‘F’. Participants were recorded doing this for 60 seconds. Then followed an abstraction task: participants were shown two sets of words, ‘train – bicycle’ and ‘watch – ruler’, and were asked to type for each set of words how they are alike. Thereafter, participants were asked to recall and type the five words they heard twice at the beginning of the screening (‘face’, ‘velvet’, ‘church’, ‘daisy’, and ‘red’). The last task tested the orientation of the participants. Here, they had to type the current year, month, day of the month, and day of the week.

2.5 Data treatment and statistical procedure

Due to a technical error, three items for each participant were missing from the data, one item from each working memory condition. This comes down to 5% of all items (N = 378 of N = 7560). These missing items included targets (N = 243) as well as fillers (N = 135).

To test whether old and young adults differed from each other and across the three working memory conditions on the performance of the sentence verification task, mixed effects modeling was used (Baayen, Davidson & Bates, 2008; Quené & Van den Bergh, 2008b). This method has several theoretical advantages over a mixed ANOVA (Quené & Van den Bergh, 2008a; Jaeger, 2008; Barr, Levy, Scheepers & Tily, 2013). Mixed effects modelling combines the ability to account for random effects coming from items and subjects with logistic regression. The binomial dependent variable in the model used in this study was the response (true or false) to the underinformative sentences, with age group (young vs. old) and working memory condition (no-load vs. low-load vs. high-load) as predictor variables. The incorporated random effects were the intercepts of item and subject.

Moreover, results on the fillers and dot task were studied descriptively. Also, a correlation analysis between the dot task and the number of pragmatic answers on the sentence verification task was carried out. This would test the predictions of the theories about effortful scalar inferencing concerning a possible trade-off between tasks. A paired samples t-test was used to statistically test the difference between the low-load condition and the high-load condition, in order to enable a direct comparison with De Neys and Schaeken's (2007) data. Lastly, an independent samples t-test was used for the Corsi block-tapping task data, in order to ensure that the two age groups significantly differed from each other with regard to working memory ability.

All statistics, including descriptive information and correlation analyses, were run in R version 4.1.0 (R Core Team, 2021). The mixed models analysis was performed with the use of the *lme4* (Bates, Maechler, Bolker & Walker, 2015) and *car* (Fox & Weisberg, 2019) packages.

3. Results

3.1 Corsi block-tapping task

The Corsi block-tapping task was administered in order to determine differences in working memory performance between the old adult group (*N* = 63) and the young adult group (*N* = 63). The mean Corsi span in the old adult group was 4.73 (*SD* = 1.88), whereas the mean Corsi span in the young adult group was 5.82 (*SD* = 1.71). An independent samples t-test showed that the difference in Corsi span between these groups was statistically significant, $t(124) = -3.42$, $p < .001$, $d = .61$, 95% CI [-1.73, -0.46].

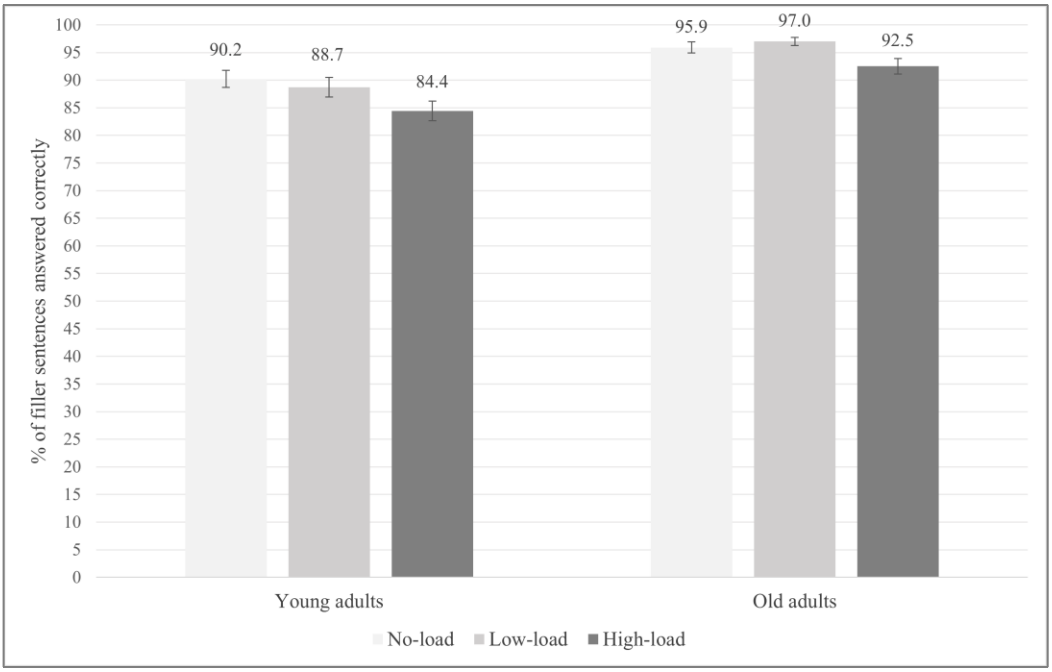
3.2 Sentence verification task

3.2.1 Filler sentences

Five types of filler sentences, which also acted as control sentences, were used in this study. These sentences were either always true (e.g., *Some insects are wasps*, *All wasps are insects*) or always false (e.g., *All insects are wasps*, *Some pigeons are insects*, *All wasps are trees*). Overall, 91.5% of all filler sentences (*N* = 3645) was answered correctly by the participants. The old adults answered a mean number of 9.51 (*SD* = 0.08) of ten fillers correctly, the young adults answered a mean number of 8.77 (*SD* = 0.13) of ten fillers correctly. Thus, the old adults performed better on the fillers (95.1% of all fillers correct) than the young adults (87.8% of all fillers correct). Performance on the fillers also differed across working memory condition. See Figure 5 below for the results of the fillers split by age group and working memory condition.

Figure 5

Mean Percentage of Filler Sentences Answered Correctly (Per Age Group and Per Working Memory Condition)



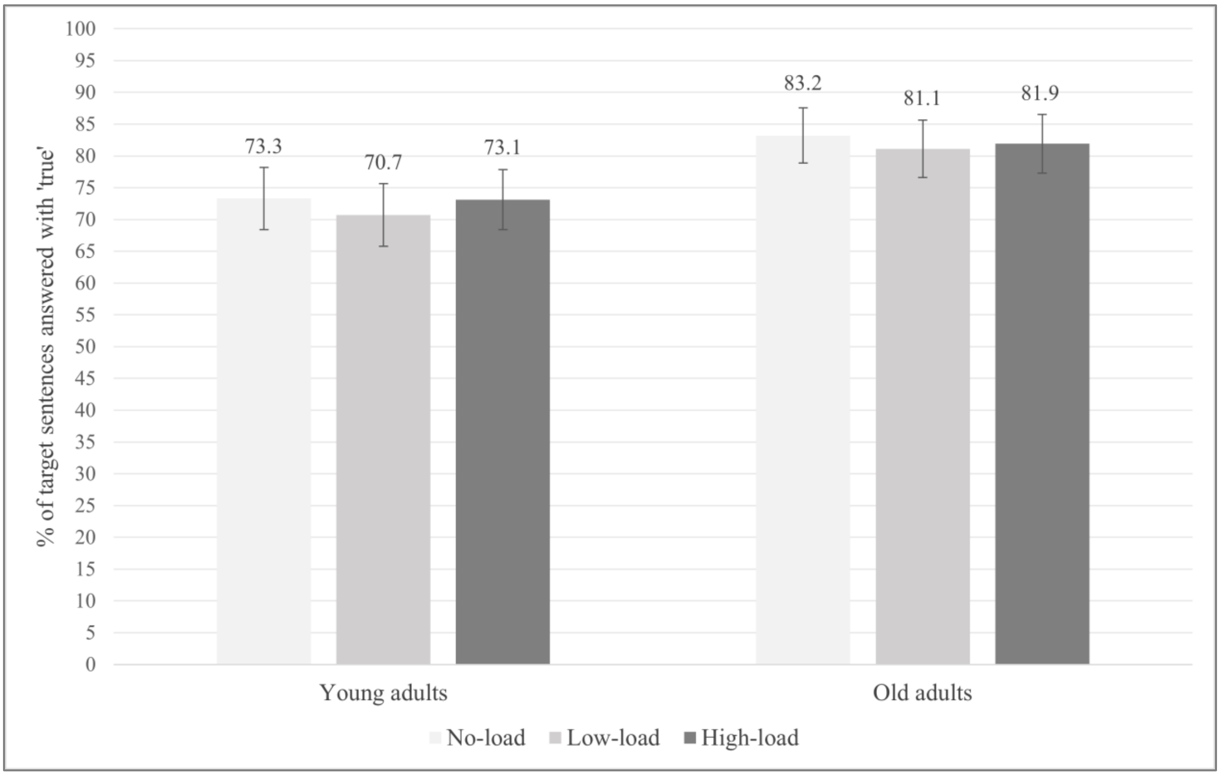
Note. Error bars reflect Standard Error.

3.2.2 Underinformative target sentences: descriptive results

Overall, the young adults gave a mean number of 7.24 (*SD* = 3.53) of ‘true’ answers to ten underinformative target sentences. This is equivalent to 72.4% of all underinformative sentences. For the old adults, this mean was 8.21 (*SD* = 3.47), i.e., 82.1% of all underinformative sentences. Figure 6 below shows the mean percentages of ‘true’ responses to the underinformative target sentences for each age group and each working memory condition. As can be seen from this figure, both age groups responded to the underinformative target sentences with ‘true’ the least in the low-load condition, but the differences between the conditions seem to be minimal.

Figure 6

Mean Percentage of Underinformative Target Sentences Answered with 'True' (Per Age Group and Per Working Memory Condition)



Note. Error bars reflect Standard Error.

In addition to looking at the percentages of target sentences answered with ‘true’, it might be informative to look at the distribution of individual proportions of how target sentences were answered. Figures 7 and 8 below represent the distribution of the participants’ answers split for age group and for working memory condition respectively. The black squares in these figures display the mean percentage for each group that is displayed. The grey areas represent the distribution of every participant’s mean proportion of ‘true’ responses to the underinformative sentences.

Figure 7

Distributions of Individual Proportions of 'True' Responses to the Underinformative Target Sentences (Per Age Group)

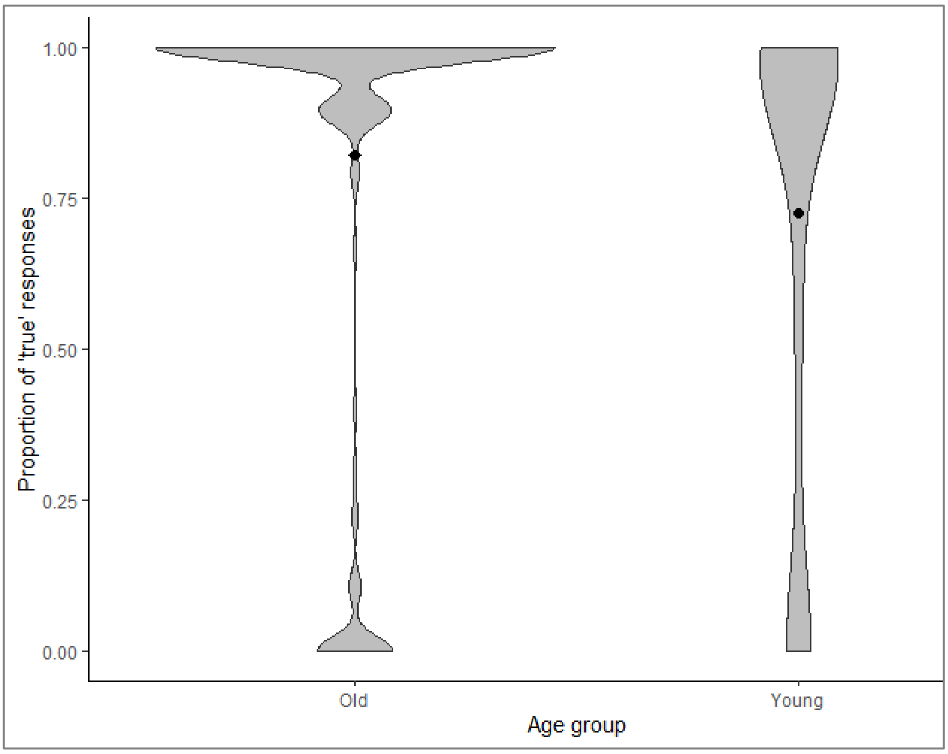
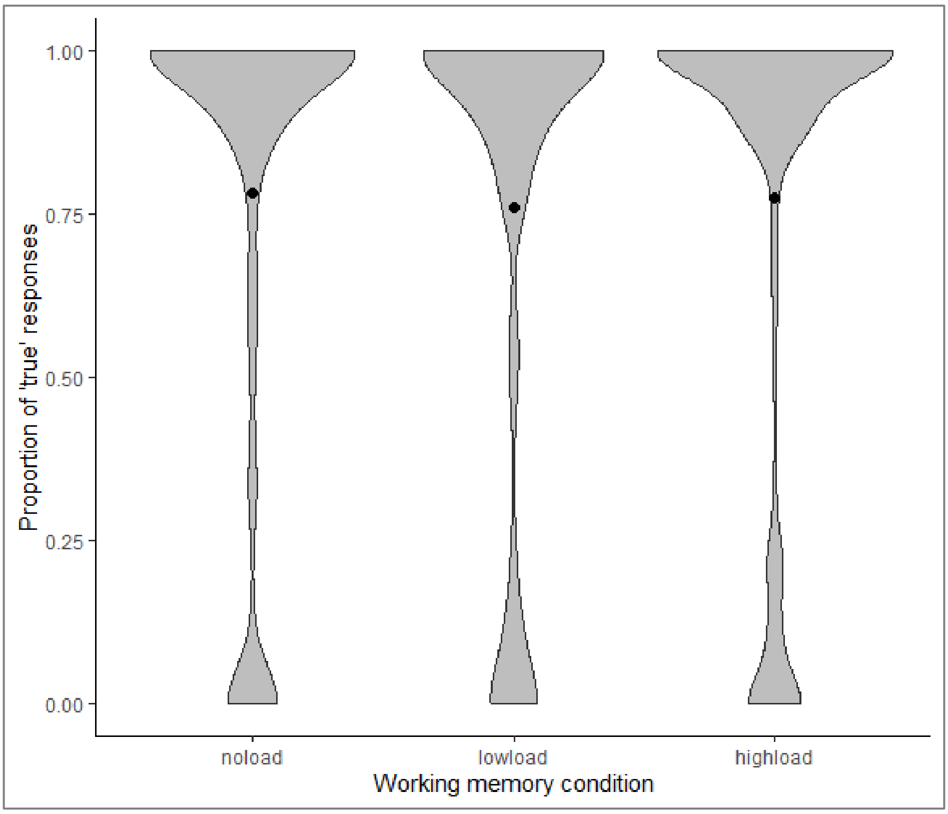


Figure 8

Distributions of Individual Proportions of 'True' Responses to the Underinformative Target Sentences (Per Working Memory Condition)



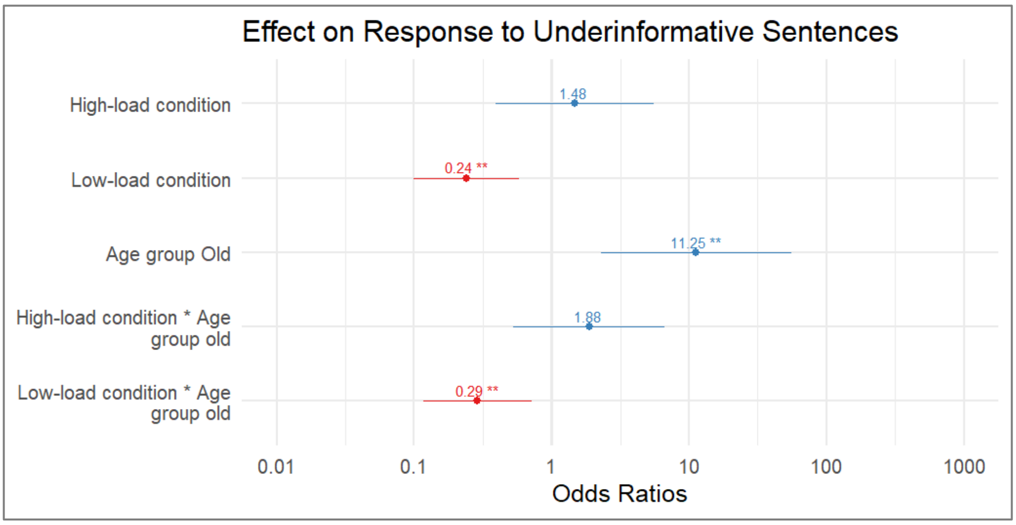
3.2.3 Underinformative target sentences: Mixed effects results

A mixed effects model was estimated, with Item and Subject as random effects, Age group and Working memory condition (and their interaction) as fixed effects, and Response to the underinformative sentences as dependent variable. This model had the maximal random effects structure, with random intercepts for Subject and Item and random slopes for the fixed effects (and their interaction) for Subject and Item. An optimizer (called 'bobyqa') was added to this model, so the model could converge better. In order to compute a p-value for the predictors in this model, a Type III Wald Chi-square test was used. This test indicates whether adding a certain variable (or interaction) significantly increases the ability of the model to explain the data. The main effect of Age group ($\chi^2(1) = 8.87$, $p = .003$), the main effect of Working memory condition ($\chi^2(2) = 10.71$, $p = .005$) and the interaction effect between Age group and Working memory condition ($\chi^2(2) = 7.16$, $p = .023$) were significant predictors for the response on the underinformative sentences. This maximal model fitted the data significantly better ($p < .001$) than a model with a less sophisticated random effects structure (consisting only of the random intercepts of the predictor variables).

Figure 9 below shows the odds ratios of the contrasts in the above-described maximal model (see Appendix 7 for details). For these contrasts, the reference levels are Age group Young, No-load condition, and the interaction between these two. The old adult group was significantly different from the young adult group (*OR* = 11.25, $p = .003$, 95% CI [2.29, 55.31]). The results also imply that the effect of Working memory condition is significant for the difference between the no-load condition and low-load condition (*OR* = 1.48, $p = .001$, 95% CI [0.39, 5.53]), but this is not the case for the difference between the no-load condition and the high-load condition (*OR* = 0.24, $p = .564$, 95% CI [0.10, 0.58]). The same holds for the interaction effect, for which only the interaction between the low-load condition and age group is significant (*OR* = 0.29, $p = .008$, 95% CI [0.12, 0.72]).

Figure 9

Odds Ratios for Each Contrast in the Maximal Mixed Effects Model



Note: ** = p < .01

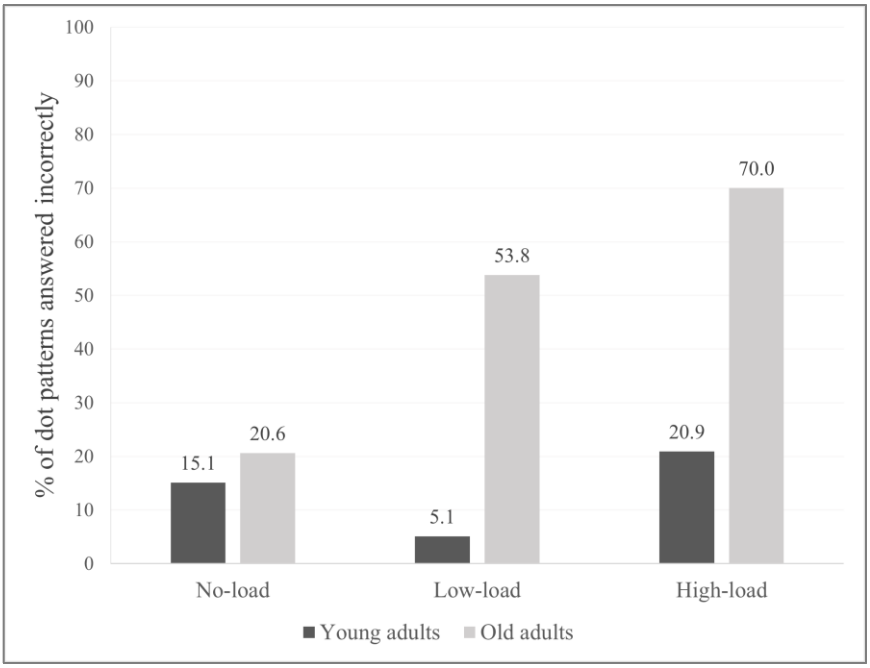
De Neys and Schaeken (2007) looked at the difference between the low-load and high-load conditions with a paired samples t-test. This was also done for the data in the current study in order to enable direct comparison. On average, the mean number of logical responses by the young adults was lower in the low-load condition (M = 7.07, SD = 3.89) than in and the high-load condition (M = 7.31, SD = 3.74). This difference, 0.25, 95%CI[-0.99, 0.49], was not significant (t (62) = -0.67, p = .507). The mean number of logical responses by the old adults was also lower in the low-load condition (M = 8.11, SD = 3.58) than in the high-load condition (M = 8.19, SD = 3.69), and the difference of 0.08, 95%CI[-0.26, 0.09] was not significant either (t(62) = -0.95, p = .345).

3.3 Dot task

The results of the dot task were analyzed in order to explore whether and where participants made mistakes in remembering the dot patterns. In the no-load condition, participants did not have to remember any pattern, which was explicitly made clear in the instructions. Participants saw an empty grid before seeing the sentences. However, they did see an answer grid after the sentences which was similar to the low-load and high-load conditions, meaning that they were able to click a pattern in this condition. The young adult group submitted 15.1% incorrect dot patterns in the no-load condition and for the old adult group this was 20.6%. These mistakes mainly consisted of participants clicking all the boxes, only the box in the middle, or a random pattern, rather than leave it empty, as they were supposed to do. In the conditions where participants actually had to recall patterns, the two age groups seemed to differ greatly in the amount of recalled incorrect patterns (see Figure 10 below).

Figure 10

Percentage of Dot Patterns that were Answered Incorrectly (Per Age Group and Per Working Memory Condition)



A correlation analysis was carried out between the number of ‘false’ answers to the underinformative target sentences (i.e., the pragmatic interpretations) and the number of correctly recalled dot patterns, in order to test the predictions of theories of scalar inferencing about a trade-off. The correlation analysis showed the following results. For the young adults, the correlation in the low-load condition was r = -0.22 (p = .003), and the correlation in the high-load condition was r = -0.28 (p = .025). Old adults showed a correlation of r = -0.025 (p = .840) in the low-load condition, and in the high-load a correlation of r = -0.006 (p = .960). The weak, but statistically significant negative correlations for the young adults indicate that the more dot patterns were recalled correctly, the more they answered logically (i.e., the less pragmatic interpretations). The minimal and insignificant correlations for the old adults suggest that the number of pragmatic answers does not correlate with the performance on the dot task.

4. Discussion

4.1 Summary of results

Overall, the results showed that the participants in this study tended to judge the underinformative target sentences more logically than pragmatically. This can be concluded from the finding that both age groups judged the underinformative sentences with 'true' more often that they did with 'false', indicating more logical interpretations of the sentences than pragmatic interpretations. Age group and working memory condition, as well as the interaction between these variables, were shown to be significant predictors of the response to underinformative sentences in a mixed model analysis that accounted for the random effects of subject and item. The descriptive data demonstrate that the old adult group, who were shown to have a significantly lower working memory ability than the young adult group, assigned the logical interpretation more often to the underinformative sentences than the young adult group did. Regarding the working memory conditions, participants seemed to have fewer pragmatic interpretations (and thus more logical interpretations) to underinformative target sentences in the no-load and high-load conditions. Only the low-load condition differed significantly from the no-load condition, not the high-load condition. This pattern was less strong for the old adults. The low-load condition and the high-load condition did not significantly differ from each other for either age group, as shown with a paired samples t-test.

The control sentences (that were also fillers in this study) were either unequivocally true or false. 91.5% of all fillers were judged correctly, but the old adult group seemed to perform better than the young adults. Moreover, both age groups seemed to answer the fillers less correctly in the high-load condition. This trend, that fillers were judged less correctly with more working memory load, was more evident for the young adults than for the old adults.

Lastly, the dot task was analyzed. The descriptive results demonstrated that the old adult group made more mistakes in recalling the patterns than the young adult group did. Both groups made more mistakes in the high-load condition than in the low-load condition. For the young adult group, a correlation analysis showed a small and statistically significant negative correlation between the number of pragmatic answers on the sentence verification task and the number of correctly recalled dot patterns on the dot task. However, the old adult group showed a smaller and statistically insignificant negative correlation.

4.2 Interpretation of results

In section 1.5.2 of this thesis, some hypotheses were proposed. I will first discuss the hypotheses that were based on different theories about how scalar terms are processed, and thus made divergent predictions regarding the effect of working memory burdening on the interpretations of the underinformative sentences with *some*. After this, I will discuss the expectations about the effect of healthy aging on the interpretation of the underinformative sentences with *some*. Then, attention will be paid to the comparison of the results of this study with those of De Neys and Schaeken (2007). Finally, the research question of this study will be answered.

4.2.1 Results in light of theories on scalar inferencing

The default approach to scalar inferencing assumes that the logical interpretation of the scalar term *some* is cognitively effortful since the pragmatic interpretation is the default. In this case, conditions in which working memory is burdened more by a dual task would result in more pragmatic interpretations and less logical interpretations. This trend is not reflected by the general results. In fact, both age groups seem to show more logical (and thus fewer pragmatic) interpretations in both the low-load condition and the high-load condition when compared to the no-load condition. Moreover, the data showed no evidence of less logical interpretations in the high-load condition than in the low-load condition.

The non-default approaches, on the other hand, pose that the pragmatic interpretation is effortful and the logical interpretation is the default. This would be reflected in more logical interpretations and less pragmatic interpretations in conditions in which working memory is burdened more. This is exactly what De Neys and Schaeken (2007), amongst others, have demonstrated in their study. The current study found that in both age groups there seem to be fewer pragmatic responses to the underinformative sentences in the high-load condition than in the low-load condition, with the difference being more substantial for the young adult group than for the old adult group. De Neys and Schaeken (2007) found a similar difference between these conditions for the young adults, which was statistically significant. In the current study, however, the difference between the low-load and high-load conditions was not statistically significant, for the young adults as well as the old adults. What is more, the fact that there were even fewer pragmatic interpretations in the no-load condition than in the other working memory load conditions for both age groups disputes the conclusion that the pragmatic interpretation would be effortful.

The correlation analysis between the number of pragmatic answers to the underinformative sentences and the number of correctly recalled dot patterns of the dot task also gives insight into the theories on scalar inferencing. The reasoning behind this is that if the pragmatic responses to the underinformative sentences are cognitively costly (non-default theory), this could take away cognitive resources from the dot task. The latter might result in less correctly recalled patterns, and this suggests, then, that there might be a trade-off between tasks. The young adults showed a small but significant negative correlation, suggesting that the fewer pragmatic inferences they made, the better they performed on the dot task. This finding is more supportive of the non-default theory on scalar inferencing than of the default theory. However, this correlation was weak by conventional standards. Furthermore, the old adult group showed only a minimal and insignificant correlation.

Finally, the default and non-default approaches indirectly make predictions about differences between age groups, provided that they differ in their working memory ability. In the case of the default approaches, a burdened working memory would mean less logical and more pragmatic interpretations, since the logical interpretation is effortful (e.g., Feeney, Scrafton & Duckworth, 2004; Holtgraves & Kraus, 2018). This would reflect itself in the old adult group making more pragmatic and less logical interpretations than the young adult group. The results of this study, though, show the opposite: the old adult group overall responded less pragmatically than the young adult group. Indeed, this finding is more in line with the prediction of the non-default approaches, which suggests that a burdened working memory ability causes fewer pragmatic interpretations and more logical interpretations (e.g., De Neys & Schaeken, 2007; Marty & Chemla, 2013; Cho, 2020).

Another, more straightforward hypothesis for the age groups was based on the study by Verschueren et al. (2006). Their study found that healthy old adults did not differ significantly in the processing and interpretation of conditional reasoning from healthy young adults, while the old adult group did have significantly lower working memory abilities than the young adult group. The authors speculated that the old adults compensated for their working memory disabilities with crystallized pragmatics, which are characterized by enhanced pragmatic skills (Verschueren et al., 2006). Nonetheless, their finding is not replicated for the scalar inferences in the current study, since the old adult group significantly differed from the young adult group in their interpretations.

In short, the results of the current study do not offer strong support for either the default or the non-default theories. This is mainly because the working memory manipulation did not seem to have an effect on either more logical or more pragmatic responses. Nevertheless, the current study did provide some evidence in line with the predictions of the non-default approach: the results of the correlation analyses between the number of pragmatic answers and the number of correctly recalled dot patterns and the results of the differences between age groups.

4.2.2 Results in light of De Neys and Schaeken (2007)

A complementary goal of this thesis was to replicate the study by De Neys and Schaeken (2007). To start with, the findings of the current study deviate notably from those of De Neys and Schaeken (2007) with regard to the total number of pragmatic responses. The current study found that the young adults judged 72.4% of all underinformative sentences with ‘true’, thus having a logical interpretation 72.4% of the time. De Neys and Schaeken (2007), however, found the exact opposite, namely that their group of young adults had a pragmatic interpretation in 76.1% of all underinformative sentences. Because the design of the current study was created to be as close to De Neys and Schaeken (2007) as possible, it is surprising why so many more participants answered logically in the current study. The survey used in the current study might give some insight in this (question 1 and 2, see section 2.3.5 and Appendix 6). A couple of participants who answered logically to the underinformative sentences presented in the survey (*Some robins are birds* and *Some ants are insects*) appeared to interpret the task as more of a general knowledge task, where the goal was to question them about their knowledge of various species. For example, if they were presented with the sentence *Some ants are insects*, they seemed to focus more on deciding whether an ant is indeed an insect, rather than on whether *some* or *all* ants are insects. However, extensive and precise qualitative research is needed to draw more generalizable conclusions.

Another difference between De Neys and Schaeken (2007) and this current study concerns the effect of the working memory manipulation. De Neys and Schaeken (2007) found that their young adult participants had a pragmatic interpretation for 78.9% of the underinformative sentences in the low-load condition, and a pragmatic interpretation for 73.2% of the underinformative sentences in the high-load condition. This difference of 5.7% was statistically significant, which led them to conclude that the participants made fewer pragmatic inferences when working memory was burdened more. The young adults in the current study showed a pragmatic interpretation for 29.3% of the underinformative sentences in the low-load condition, and for 26.9% in the high-load condition. This difference of 2.4% was less than De Neys and Schaeken (2007) found and not significant. Perhaps more importantly, De Neys and Schaeken's (2007) conclusion that the working memory manipulation resulted in less pragmatic interpretations for the underinformative sentences is called into question by the findings from the no-load condition in the current study. That is, the results of the current study showed that the number of pragmatic interpretations was the lowest in the no-load condition, in which working memory is not burdened at all, compared to the other two conditions.

Other differences with De Neys and Schaeken's (2007) data were that the healthy young adults in the current study made more errors in the performance on the dot task and answered the filler sentences incorrectly more often. One difference with De Neys and Schaeken's (2007) experimental design that could have caused the worse performance on the dot task is the absence of feedback the participants received in the current study. In De Neys and Schaeken's (2007) experiment, participants received feedback if they recalled a dot pattern incorrectly, which emphasized that they should recall it correctly in the next trials. This feature was not present in the design of the current study due to technical limitations. Participants in the current study received no feedback about whether they recalled the patterns (in)correctly. This might have caused participants to focus less on recalling the patterns correctly. Another difference that might have caused both a worse performance on the dot task and a worse performance on the filler sentences, is the fact that the current study did not take place in a research lab. Rather, participants performed the experiment online from their private computer, without supervision. Maybe participants from De Neys and Schaeken (2007), who did the experiment in a supervised lab, felt more pressured to perform both tasks well than did the participants in the current study.

4.2.3 Results in light of the research question

The research question of this study was as follows: to what extent do healthy old adults differ from healthy young adults in their interpretations of the scalar term *some*, and what is the effect of a working memory load on this difference? The results have shown that, overall, the healthy old adults assigned logical interpretations significantly more often when judging underinformative sentences with the scalar term some than the healthy young adults did. The mixed model analysis indicates that the effect of age played a larger role in some working memory conditions than in others, but there does not seem to be a clear and strong pattern here. Moreover, there appeared to be no clear effect of working memory manipulation, contrary to the findings of De Neys and Schaeken (2007). Based on these results, there seems to be a contradiction. On the one hand, working memory manipulation seems to have no effect on the interpretations of the underinformative sentences. On the other hand, there does seem to be a difference in interpretations between old adults and young adults, who significantly differ in their working memory abilities (as shown by the Corsi block-tapping task). The alternative explanation in accordance with Verschueren et al. (2006), which was that old adults compensate for their decreased working memory abilities, does not seem to fit with the findings either. Contrary to their expectations, the current study did find differences between the young and old adults in the amount of pragmatic or logical responses.

How can the results of the current study be explained then? One possibility is that healthy old adults (possibly unknowingly) employed a different strategy from healthy young adults when performing the cognitive task. Looking at Figure 7 (on page 26), it is clear that the distribution of interpretations is more concentrated at the extremities of the Y-axis for the old adult group than for the young adult group. The young adult group shows a wider distribution of the proportion of 'true' responses to the underinformative sentences. This could indicate that the old adults in general might have had a different strategy for the whole task, than did the young adults. Literature about this topic has generally shown that young and old adults significantly differ in strategy when performing cognitive tasks (e.g., Lemaire, 2010; Hodzik & Lemaire, 2011; Barulli, Rakitin, Lemaire & Stern, 2013; Roquet & Lemaire, 2019).

The next logical question then becomes what this strategy might have been. Looking at the results, it could be argued that the old adults focused more on the sentence verification task (hence the concentrated distribution of interpretations) and less on the patterns of the dot task. The latter can be seen in Figure 10 (on page 28), which shows that the old adults made considerably more mistakes on the dot task. Further evidence from the current study that the old adults might have focused more on the sentence verification task comes from the results of the filler sentences. Figure 5 (on page 24) shows that the old adults answered the fillers correctly more often than the young adults. The young adults, in turn, seem to have focused more on the dual task as a whole. This can be demonstrated with the fact that the young adults performed better on the dot task but worse on the fillers in the sentence verification task. The idea that strategy might have played a role in the current study can be supported by studies that examine age-related differences on dual tasks, which generally show a divergence in approaches to dual tasks between young and old adults (e.g., Hein & Schubert, 2004; Holtzer, Stern & Rakitin, 2005; Göthe, Oberauer & Kliegl, 2007; Brustio, Magistro, Zecca, Rabaglietti & Luibicich, 2017).

4.3 Limitations and confounds

The results of this study and the conclusions drawn from them can be challenged by several limitations and/or confounds of this study.

As mentioned before, this study recruited participants online via Prolific (www.prolific.co). It might be possible that a population of people with certain kinds of characteristics are more active on this website, which could have influenced the results. For example, a study by Khorsheed, Rashid, Nimeschisalem, Imm and Price (2021) found that people with certain autistic-like traits tended to interpret scalar terms more logically than people without these traits. It is a possibility that participants with certain personality and /or autistic traits were overrepresented in the current study, leading to a far higher number of logical interpretations when compared to the results of De Neys and Schaeken (2007).

Another confounding influence in this study could have had to do with the stimuli (see Appendix 1). It might have been the case that the 'members' (e.g., *bees*) of the 'categories' (e.g., *insects*) in the stimuli sentences were not familiar words for the participants in the current study. There are two possible causes of this. First of all, the current study copied all sentences for the low-load condition and the high-load condition from De Neys and Schaeken (2007). Since an extra condition was added (the no-load), more sentences had to be added, but the five categories remained the same (birds, flowers, insects, fish, and trees). Because of this, the number of members per category increased, which increased the probability that certain members were less close to the prototype of the category, and thus less familiar to participants. The second possible cause is that the participants resided in various places around the world at the time of the study (see Figure 1 on page 16) An example of this is that some participants mentioned in the survey that the word *ladybug* is not as common in the UK, and that the word *ladybird* is used more often in that country. There were even a couple of participants from the UK that did not recognize the word *ladybug* at all. This shows that there could have been differences between participants in their answers to the stimuli sentences purely based on the region where they reside. All of this can be supported with evidence from the survey. That is, participants often indicated in the survey that they were not familiar with one or more words used in the experiment (question 3 of the survey, see section 2.3.5 and Appendix 6).

Another limitation of the experiment in this study is the presentation of the dot task in the no-load condition. In that condition, participants saw an empty grid without patterns, because they did not have to recall one. Quite some participants indicated in the survey that it was unclear to them that they did not have to recall a pattern in this condition (although this was clearly explained in the institutions, and participants also performed a test item for the no-load condition where this was made clear). This can also be seen in Figure 10, as the old adults and the young adults submitted checked boxes in the no-load condition, which is registered as an incorrectly recalled pattern (15.1% of all patters in the no-load condition for the young adults, 20.6% of all patters in the no-load condition for the old adults). The fact that the dot task in the no-load condition was unclear to participants might have caused precisely the opposite of what was the intention of the no-load condition: make it effortful.

4.4 Recommendations for future research

In order to follow up on the current study, future research could replicate this and De Neys and Schaeken's (2007) study with more participants in order to obtain more statistical power. With this, more insight will also be gained in how healthy aging plays a role in scalar inferencing. Moreover, one could also dive deeper into how dual task strategies influence the interpretation of scalar terms, for example by employing a thinking out loud experiment or choosing a task other than the dot task. A thinking-out-loud experiment could, for example, uncover why participants choose a certain interpretation, which might reveal several trends in this regard. Choosing a different task to be the second task alongside a sentence verification task might give insight into whether this other task, be it representative of working memory or another cognitive modality, has similar effects on the strategies that participants use.

Finally, the current study recruited monolingual as well as bilingual participants, but the effect of this on the interpretations of the scalar term *some* in their first language was not explored. However, recent research where scalar inferencing is studied in bilinguals indeed found effects of bilingualism. Mazzaggio, Panizza and Surian (2021), for example, looked at scalar inferencing by participants with English as a second language. They found that participants were less likely to interpret underinformative sentences pragmatically in their second language than in their first language. Khorsheed et al. (2021) even found that the interpretation of scalar terms in the second language differs for participants with a weaker second language proficiency than participants with a proficiency advantage. Regarding the current study, it would be interesting to know whether a difference like this also exists between monolingual and bilingual participants when testing scalar terms in the first language, since almost half of the participants in the current study spoke more than one language (N = 55).

5. Conclusion

This study has researched the processing of the scalar term *some* by healthy old and healthy young adults. This was done by testing so-called underinformative sentences, which have two interpretations: a logical interpretation and a pragmatic interpretation. Many studies have shown that the pragmatic interpretation of the scalar term *some* is cognitively effortful. The goal of the current study was to replicate these findings and additionally examine whether the healthy old adult group, which was characterized by a decreased working memory ability, would make fewer pragmatic interpretations than the young adult group. Participants were presented with underinformative sentences which they had to judge with true or false. Alongside this sentence verification task, participants also performed a dot task in order to burden their working memory by having to recall patterns. There were three conditions: the no-load condition (in which the participants did not have to recall any patterns), the low-load condition (in which the participants had to recall a simple pattern), and the high-load condition (in which the participants had to recall a complex pattern).

The specific research question in this study was as follows: to what extent do healthy old adults differ from healthy young adults in their interpretations of the scalar term *some*, and what is the effect of a working memory manipulation on this difference? First of all, the results showed that the old adult group made significantly more logical interpretations than did the young adult group. Secondly, there did not seem to be an effect of working memory manipulation on the type of interpretation (logical or pragmatic) of the scalar term *some*, for neither age group. The results of the working memory manipulation are puzzling in light of the existing theories about the processing of scalar terms. These theories predict that the more working memory is burdened, there should be either more logical (non-default approaches) or more pragmatic interpretations (default approaches). Neither is the case in the current study. The absence of an effect of working memory manipulation means that the difference between the young adult group and old adult group cannot be explained by this. Moreover, the fact that this study found no effect of working memory manipulation suggests that a decreased working memory ability cannot be the cause of the old adult group making more logical interpretations in general than the young adult group. In this study I propose an alternative explanation to the above-mentioned difference between age groups, namely that there is an effect of age on general (dual) task performance.

So, coming back to Mr. Bean and Irma from the introduction. What might Irma have expected when Mr. Bean handed her the popcorn bucket while saying 'I ate some of it'? If she were to be an old lady, it might have been more likely that she already expected Mr. Bean to have eaten all of the popcorn, as the results of this study show. However, being Mr. Bean's girlfriend, Irma did not necessarily have to be of age to expect this from him. After all, it is in Mr. Bean's nature to fool people, especially his girlfriend.

References

Albinet, C. T., Boucard, G., Bouquet, C. A., & Audiffren, M. (2012). Processing Speed and Executive Functions in Cognitive Aging: How to Disentangle their Mutual Relationship? *Brain and Cognition, 79*(1), 1-11. doi: 10.1016/j.bandc.2012.02.001

Audacity Team (2021). Audacity®: Free Audio Editor and Recorder [Computer application]. Version 3.0.2 retrieved May 18th 2021 from https://audacityteam.org/

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal and Memory and Language 59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Baddeley, A. D. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience, 4,* 829-838. doi: 10.1038/nrn1201

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. doi: 10.1016/j.jml.2012.11.001

Barulli, D. J., Rakitin, B. C., Lemaire, P., & Stern, Y. (2013). The Influence of Cognitive Reserve on Strategy Selection in Normal Aging. *Journal of the International Neuropsychological Society : JINS, 19*(7), 841-844. doi: 10.1017/S1355617713000593

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi Block-Tapping Task: Methodological and Theoretical Considerations. *Brain and Cognition*, *38*(3), 317-338. doi: 10.1006/brcg.1998.1039

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental Rotation: Effects of Stimulus Complexity and Familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 12-23. doi: 10.1037/0096-1523.14.1.12

Bielak, A. A. M., Hultsch, D. F., Kadlec, H., & Strauss, E. (2007). Was that Part of the Story or Did I Just Think So? Age and Cognitive Status Differences in Inference and Story Recognition. *Experimental Aging Research, 33*(3), 295-322. doi: 10.1080/03610730701319103

Bonnefon, J. F., De Neys, W., & Feeney, A. (2011). Processing Scalar Inferences in Face- Threatening Contexts. In C. H. L. Carlson, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society.* Austin, Texas: Cognitive Science Society.

Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When Some is Actually All: Scalar Inferences in Face-Threatening Contexts. *Cognition, 112*(2), 249-258. doi: 10.1016/j.cognition.2009.05.005

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing Speed from Accuracy in Scalar Implicatures. *Journal of Memory and Language, 66*(1), 123-142. doi: 10.1016/j.jml.2011.09.005

Bott, L., & Noveck, I. A. (2004). Some Utterances are Underinformative: The Onset and Time Course of Scalar Inferences. *Journal of Memory and Language, 51*(3), 437-457. doi: 10.1016/j.jml.2004.05.006

Breheny, R., Katsos, N., & Williams, J. (2006). Are Generalised Scalar Implicatures Generated by Default? An On-line Investigation into the Role of Context in Generating Pragmatic Inferences. *Cognition, 100*(3), 434-463. doi: 10.1016/j.cognition.2005.07.003

Brown, L. A. (2016). Spatial-Sequential Working Memory in Younger and Older Adults: Age Predicts Backward Recall Performance within Both Age Groups. *Frontiers in Psychology,* 7:1514, 1-11. doi: 10.3389/fpsyg.2016.01514

Brown, L. J. E., Ferner, H. S., Robertson, J., Mills, N. L., Pessotto, R., Deary, I. J., & MacLullich, A. M. J. (2011). Differential Effects of Delirium on Fluid and Crystallized Cognitive Abilities. *Archives of Gerontology and Geriatrics, 52*, 153-158. doi: 10.1016/j.archger.2010.03.005

Brickman, A. M., & Stern, Y. (2009). Aging and Memory in Humans. *Encyclopedia of Neuroscience, 1,* 175-180. doi: 10.1016/B978-008045046-9.00745-2

Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: Implementation and Testing of the Corsi Block-Tapping Task for Digital Tablets. *Frontiers in Psychology,* 5:939, 1-8. doi: 10.3389/fpsyg.2014.00939

Brustio, P. R., Magistro, D., Zecca, M., Rabaglietti, E., & Liubicich, M. E. (2017). Age-Related Decrements in Dual-Task Performance: Comparison of Different Mobility and Cognitive Tasks. A Cross Sectional Study. *PLOS ONE, 12*(7), 1-15. doi: 10.1371/journal.pone.0181698

Carston, R. (1998). Informativeness, relevance, and scalar implicature. In R. Carston, & S. Uchida (Eds.), *Relevance theory: Applications and Implications* (pp. 179-238). Amsterdam: John Benjamins.

Chemla, E., & Singh, R. (2014). Remarks on the Experimental Turn in the Study of Scalar Implicature, Part I. *Language and Linguistics Compass, 8*(9), 373-386. doi: 10.1111/lnc3.12081

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond* (pp. 39–103). Oxford: Oxford University Press.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar Implicature as a Grammatical Phenomenon. In K. van Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2297-2331). Berlin: Mouton de Gruyter.

Cho, J. (2020). Memory Load Effect in the Real-Time Processing of Scalar Implicatures. *Journal of Psycholinguistic Research, 49*, 865-884. doi: 10.1007/s10936-020-09726-3

Cohen, G. (1981). Inferential Reasoning in Old Age. *Cognition*, *9*(1), 59-72. doi: 10.1016/0010-0277(81)90014-7

Corsi, P. M. (1972). *Human Memory and the Medial Temporal Region of the Brain.* Doctoral Thesis at McGill University (Canada). Retrieved from https://escholarship.mcgill.ca/concern/theses/05741s554

D’Antuono, G., Miani, M., Marin, D., Boccia, M., & Piccardi, L. (2020). Effect of Ageing on Verbal and Visuo-Spatial Working Memory: Evidence from 880 Individuals. *Applied Neuropsychology: Adult.* doi: 10.1080/23279095.2020.1789986

Degen, J., Reeder, P., Carbary, K., & Tanenhaus, M. K. (2009). Using a Novel Experimental Paradigm to Investigate the Processing of Scalar Implicatures. *In Paper presented at the 3rd biennial meeting of experimental pragmatics.* Lyon: France.

Degen, J., & Tanenhaus, M. K. (2011). Making Inferences; the Case of Scalar Implicature Processing. In L. Carlson C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3299-3304). Austin, Texas: Cognitive Science Society.

Degen, J., & Tanenhaus, M. K. (2014). Processing Scalar Implicature: A Constraint-Based Approach. *Cognitive Science, 39*, 667-710. doi: 10.1111/cogs.12171

De Neys, W., & Schaeken, W. (2007). When People Are More Logical Under Cognitive Load: Dual Task Impact on Scalar Implicature. *Experimental Psychology, 54*, 128-133. doi: 10.1027/1618-3169.54.2.128

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some Effort for Some: Further Evidence that Scalar Implicatures are Effortful. *Quarterly Journal of Experimental Psychology, 64*(12), 2352-2367. doi: 10.1080/17470218.2011.588799

Fisk, J. E., & Sharp, C. A. (2004). Age-Related Impairment in Executive Functioning: Updating, Inhibition, Shifting, and Access. *Journal of Clinical and Experimental Neuropsychology, 27*(7), 874-890. doi: 10.1080/13803390490510680

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.

Gazes, Y., Lee, S., Sakhardande, J., Mensing, A., Razlighi, Q., Ohkawa, A., … Habeck, C. (2020). fMRI-guided White Matter Connectivity in Fluid and Crystallized Cognitive Abilities in Healthy Adults. *NeuroImage, 215*, 1-10. doi: 10.1016/j.neuroimage.2020.116809

Göthe, K., Oberauer, K., & Kliegl, R. (2007). Age Differences in Dual-Task Performance after Practice. *Psychology and Aging, 22*(3), 596-606. doi: 10.1037/0882-7974.22.3.596

Grice, H. P. (1975). Logic and Conversation. In: P. Cole & J. L. Morgan (Eds.), *Studies in syntax. Speech acts* (Vol. 3, pp. 41-58). New York, NY: Academic Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and Possibly All, Scalar Inferences are not Delayed: Evidence for Immediate Pragmatic Enrichment. *Cognition, 116*(1), 42-55. doi: 10.1016/j.cognition.2010.03.014

Hamm, V. P., & Hasher, L. (1992). Age and the Availability of Inferences. *Psychology and Aging, 7*(1), 56-64.

Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal Cognitive Aging. *Clinics in Geriatric Medicine, 29*(4), 737-752. doi: 10.1016/j.cger.2013.07.002

Head, D., Raz, N., Gunning-Dixon, F., Williamson, A., & Acker, J.D. (2002). Age-Related Differences in the Course of Cognitive Skill Acquisition: The Role of Regional Cortical Shrinkage and Cognitive Resources. *Psychology and Aging, 17*(1), 72-84. doi: 10.1037/0882-7974.17.1.72

Hein, G., & Schubert, T. (2004). Aging and Input Processing in Dual-Task Situations. *Psychology and Aging, 19*(3), 416-432. doi: 10.1037/0882-7974.19.3.416

Hodzik, S., & Lemaire, P. (2011). Inhibition and Shifting Capacities Mediate Adults' Age-Related Differences in Strategy Selection and Repertoire. *Acta Psychologica, 137*(3), 335-344. doi: 10.1016/j.actpsy.2011.04.002

Holtgraves, T., & Kraus, B. (2018). Processing Scalar Implicatures in Conversational Contexts: An ERP study. *Journal of Neurolinguistics, 46*, 93-108. doi: 10.1016/j.jneuroling.2017.12.008

Holtzer, R., Stern, Y., & Rakitin, B. C. (2005). Predicting Age-Related Dual-Task Effects With Individual Differences on Neuropsychological Tests. *Neuropsychology, 19*(1), 18-27. doi: 10.1037/0894-4105.19.1.18

Horn, L. R. (1972). On the Semantic Properties of Logical Operators in English, Unpublished doctoral dissertation, University of California, Los Angeles (Distributed by Indiana University Linguistics Club, 1976).

Horn, L. R. (2004). Implicature. In L. Horn & G. Ward (Eds), *The Handbook of Pragmatics Oxford* (pp. 3-28). Oxford: Blackwell Publishing.

Huang, Y. T., & Snedeker, J. (2009). Online Interpretation of Scalar Quantifiers: Insight into the Semantics-pragmatics Interface. *Cognitive Psychology, 58*, 376-415. doi: 10.1016/j.cogpsych.2008.09.001

Huang, Y. T., & Snedeker, J. (2011). Logic and Conversation Revisited: Evidence for a Division between Semantic and Pragmatic Content in Real-Time Language Comprehension. *Language and Cognitive Processes, 26*(8), 1161-1172. doi: 10.1080/01690965.2010.508641

Ichikawa, S. I. (1983). Verbal Memory Span, Visual Memory Span, and their Correlations with Cognitive Tasks. *Japanese Psychological Research, 25*(4), 173-180. doi: 10.4992/psycholres1954.25.173

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi: 10.1016/j.jml.2007.11.007

Janssens, L., & Schaeken, W. (2016). ‘But’ Implicatures: A Study of the Effect of Working Memory and Argument Characteristics. *Frontiers in Psychology*, 7:1520, p. 1-13. doi: 10.3389/fpsyg.2016.01520

Jarjat, G., Portrat, S., & Hot, P. (2019). Aging Influences the Efficiency of Attentional Maintenance in Verbal Working Memory. *Journals of Gerontology: Psychological Sciences, 74*(4), 600-608. doi: 10.1093/geronb/gby067

Kemtes, K. A., & Kemper, S. (1999). Aging and Resolution of Quantifier Scope Effects. *The Journals of Gerontology: Series B, 54B*(6), P-350-P360. doi: 10.1093/geronb/54B.6.P350

Kessels, R. P. C., Van Den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The Backward Span of the Corsi Block-Tapping Task and Its Association With the WAIS-III Digit Span. *Assessment, 15*(4), 426-434. doi: 10.1177/1073191108315611

Kessels, R. P. C., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., De Haan, E. H. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology, 7*(4), 252-258. doi: 10.1207/S15324826AN0704_8

Khorsheed, A., Rashid, S. M., Minehchisalem, V., Imm, L. G., & Price, J. (2021). On the Variability of Scalar Implicature Computation: Evidence from L2 English Adults. doi: 10.31234/osf.io/5gwp7

Klencklen, G., Lavenex, P. B., Bradner, C., & Lavenex, P. (2017). Working Memory Decline in Normal Aging: Memory Load and Representational Demands Affect Performance. *Learning and Motivation, 60*, 10-22. doi: 10.1016/j.lmot.2017.09.002

Kyllonen, P. C., & Cristal, R. E. (1990). Reasoning Ability is Little More than Working Memory? *Intelligence*, *14*, 389-433. doi: 10.1016/S0160-2896(05)80012-1

Lemaire, P. (2010). Cognitive Strategy Variations During Aging. *Current Direction in Psychological Science, 19*(6), 363-369. doi: 10.1177/0963721410390354

Levinson, S. (2000). *Presumptive meanings.* Cambridge, MA: MIT Press.

Lubitz, A. F., Niedeggen, M., & Feser, M. (2017). Aging and Working Memory Performance: Electrophysiological Correlates of High and Low Performing Elderly. *Neuropsychologia*, *106*, 42-51. doi: 10.1016/j.neuropsychologia.2017.09.002

Marty, P. P., & Chemla, E. (2013). Scalar Implicatures: Working Memory and a Comparison with Only. *Frontiers in Psychology,* 4:403, 1-12. doi: 10.3389/fpsyg.2013.00403

Marty, P. P., Chemla, E., & Spector, B. (2013). Interpreting Numerals and Scalar Items under Memory Load. *Lingua, 133*, 152-163. doi: 10.1016/j.lingua.2013.03.006

Mazzaggio, G., Panizza, D., & Surian, L. (2021). On the Interpretation of Scalar Implicatures in First and Second Language. *Journal of Pragmatics, 171*, 62-75. doi: 10.1016/j.pragma.2020.10.005

Meeter, M., & Hendriks, M. (2020). Geheugen. In R. Kessels, P. Eling, R. Ponds, J. Spikman & M. Van Zandvoort (Eds.), *Klinische Neuropsychologie* (pp. 197-218). Amsterdam: Boom uitgevers.

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

Nasreddine, Z. S., Phillips N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., … & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCa: a Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. doi: 10.1111/j.1532-5415.2005.53221.x

Noveck, I. A., & Posada, A. (2003). Characterizing the Time Course of an Implicature: An Evoked Potentials Study. *Brain and Language, 85*(2), 203-210. doi: 10.1016/S0093-934X(03)00053-1

Politzer-Ahles, S., & Fiorentino, R. (2013). The Realization of Scalar Inferences: Context Sensitivity without Processing Cost. *PLoS ONE, 8*(5): e63943, 1-6. doi: 10.1371/journal.pone.0063943

Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of Prefrontal Cortex in Scalar Implicatures: Evidence from Magnetoencephalography. *Language, Cognition and Neuroscience 30*(7), 853-866. doi: 10.1080/23273798.2015.1027235

Power, A. (2017). *Assessing Working Memory through Digit Span and Corsi Block Tests.* doi: 10.13140/RG.2.2.11742.38721

Quené, H., & Van den Bergh, H. (2008a). Examples of Mixed-Effects Modeling with Crossed Random Effects and with Binomial Data. *Journal of Memory and Language, 59*(4), 413-425. doi: 10.1016/j.jml.2008.02.002

Quené, H., & Van den Bergh, H. (2008b). On Multi-Level Modeling of Data from Repeated Measures Designs: A Tutorial. *Speech Communication, 43*(1-2), 103-121. doi: 10.1016/j.specom.2004.02.004

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Version 4.1.0 retrieved May 28th 2021 from https://www.R-project.org/

Reuter-Lorenz, P. A., & Sylvester, C.-Y. C. (2005). The Cognitive Neuroscience of Working Memory and Aging. In R. Cabeza, L. Nyberg, & D. Park (Eds.), *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging* (pp. 186-217). Oxford: Oxford University Press.

Roquet, A., & Lemaire, P. (2019). Strategy Variability in Numerosity Comparison Task: A Study in Young and Older Adults. *Open Psychology, 1*(1), 152-167. doi: 10.1515/psych-2018-0011

Salthouse, T. A. (1990). Working Memory as a Processing Resource in Cognitive Aging. *Developmental Review, 10*(1), 101-124. doi: 10.1016/0273-2297(90)90006-P

Salthouse, T. A. (2001). Structural Models of the Relations between Age and Measures of Cognitive Functioning. *Intelligence, 29*(2), 93-115. doi: 10.1016/S0160-2896(00)00040-4

Salthouse, T. A. (2004). What and When of Cognitive Aging. Current Directions in Psychological Science, 13(4), 140-144. doi: 10.1111/j.0963-7214.2004.00293.x

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of Adult Age and Working Memory on Reasoning and Spatial Abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition 15*(3), 507-516.

Sperber, D., & Wilson, D. (1985/1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (2002). Relevance Theory. In G. Ward, & L. Horn (Eds.), *Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.

Stoet, G. (2010). PsyToolkit – A Software Package for Programming Psychological Experiments using Linux. *Behavior Research Methods, 42*(2), 1096-1104. doi: 10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology, 44*(1), 24-31. doi: 10.1177/0098628316677643

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly All of That and Then Some: Scalar Implicatures are Understood in Two Steps. *Journal of Memory and Language, 69*(1), 18-35. doi: 10.1016/j.jml.2013.02.003

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Ageing 2019: Highlights* (ST/ESA/SER.A/430).

Van Tiel, B., Pankratz, E., & Sun, S. (2019). Scales and Scalarity: Processing Scalar Inferences. *Journal of Memory and Language, 105*, 93-107. doi: 10.1016/j.jml.2018.12.002

Van Tiel, B., Pankratz, E., Marty, P., & Sun, C. (2019). Scalar Inferences and Cognitive Load. In: M. Teresa Espinal et al. (eds.) *Proceedings of Sinn und Bedeutung 23*(2), 427-441. Universitat Autonoma de Barcelona, Bellaterra (Cerdanyola del Vallès).

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar Diversity. *Journal of Semantics, 0,* 1-39. doi: 10.1093/jos/ffu017

Van Tiel, B., Schaeken, W. (2017). Processing Coversational Implicatures: Alternatives and Counterfactual Reasoning. *Science, 41*(S5), 1119-1154. doi: 10.1111/cogs.12362

Verschueren, N., Schaeken, W., & d'Ydewalle G. (2004). Everyday Conditional Reasoning with Working Memory Preload. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society* (pp. 1399-1404). Mahwah, NJ: Erlbaum.

Verschueren, N., Schaeken, W., & Verbrugge, S. (2006). Conversational Implicatures in Counterexample Retrieval: Working Memory and Crystallized Pragmatics. *Proceedings of the 28th Annual Conference of the Cognitive Society* (pp. 2311-2316). Mawah, NJ: Erlbaum.

Wittich, W., Phillips, N., Nasreddine, Z. S., & Chertkow, H. (2010). Sensitivity and Specificity of the Montreal Cognitive Assessment Modified for Individuals Who Are Visually Impaired. *Journal of Visual Impairment & Blindness, 104*(6), 360-368. doi: 10.1177/0145482X1010400606

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). doi: 10.17605/OSF.IO/MD832

Appendix 1: Stimuli of the sentence verification task

	Set 1	*Set 2*	*Set 3*
Target sentences (underinformative)	Some eels are fish. Some carp are fish. Some oaks are trees. Some beeches are trees. Some sparrows are birds. Some robins are birds. Some flies are insects. Some mosquitoes are insects. Some roses are flowers. Some tulips are flowers.	Some ants are insects. Some bees are insects. Some canaries are birds. Some blackbirds are birds. Some daisies are flowers. Some lilies are flowers. Some firs are trees. Some birches are trees. Some trout are fish. Some sharks are fish.	Some sunflowers are flowers. Some orchids are flowers. Some ladybugs are insects. Some roaches are insects. Some salmon are fish. Some tuna are fish. Some seagulls are birds. Some doves are birds. Some pines are trees. Some chestnut trees are trees.
Filler sentences	Some birds are magpies. (true) Some insects are wasps. (true) Some pigeons are insects. (false) Some beetles are flowers. (false) All Chrysanthemum are flowers. (true) All hazels are trees. (true) All trees are elms. (false) All fish are herrings. (false) All daffodils are trees. (false) All sycamores are fish. (false)	Some flowers are carnations. (true) Some trees are willows. (true) Some crocuses are trees. (false) Some poplars are fish. (false) All cod are fish. (true) All parrots are birds. (true) All birds are crows. (false) All insects are worms. (false) All pike are birds. (false) All swallows are insects. (false)	Some trees are cedar trees. (true) Some fish are goldfish. (true) Some palm trees are fish. (false) Some flounders are insects. (false) All moths are insects. (true) All hyacinths are flowers. (true) All flowers are pansies. (false) All birds are hummingbirds. (false) All crickets are flowers. (false) All poppies are birds. (false)

Appendix 2: Patterns of the dot task

Given a 3 by 3 grid with the following boxes that can be filled with a check (indicated by the numbers 1 up to 9):

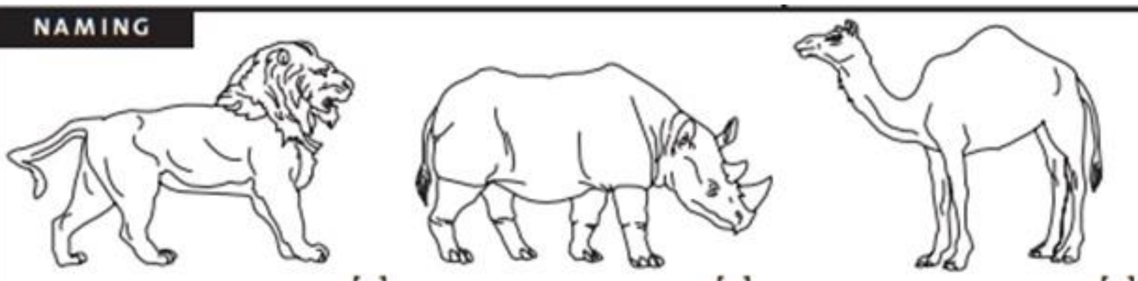
1	2	3
4	5	6
7	8	9

The following table shows the patterns used in the low-load condition and the high-load condition, with the order they were presented in. The numbers of the patterns represent the boxes checked in the grid above.

Order	Low-load pattern	High-load pattern
1.	789	1267
2.	456	3489
3.	123	2349
4.	456	3468
5.	123	3479
6.	789	1348
7.	123	1367
8.	456	2348
9.	123	1679
10.	147	2467
11.	369	3489
12.	258	1267
13.	147	1349
14.	369	3468
15.	147	1348
16.	258	2349
17.	369	3479
18.	147	3589
19.	369	2348
20.	258	1267

Appendix 3: MoCa test and scoring (as adjusted for the current study)

NAMING



[ ] [ ] [ ] __/3

MEMORY	Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.		FACE	VELVET	CHURCH	DAISY	RED
		1st trial					
		2nd trial					

No points

ATTENTION Read list of digits (1 digit/sec.). Subject has to repeat them in the forward order [ ] 2 1 8 5 4

Subject has to repeat them in the backward order [ ] 7 4 2 __/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors

[ ] FBACMNAAJKLBAFAKDEAAAJAMOFAAB __/1

Serial 7 subtraction starting at 100 [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65

4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt** __/3

LANGUAGE Repeat : I only know that John is the one to help today. [ ]

The cat always hid under the couch when dogs were in the room. [ ] __/2

Fluency / Name maximum number of words in one minute that begin with the letter F [ ] ____ (N ≥ 11 words) __/1

ABSTRACTION Similarity between e.g. banana - orange = fruit [ ] train – bicycle [ ] watch - ruler __/2

DELAYED RECALL	Has to recall words WITH NO CUE	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only	__/5
Optional	Category cue							
	Multiple choice cue							

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day __/4

© Z.Nasreddine MD Version November 7, 2004

www.mocatest.org

TOTAL __/23

Add 1 point if ≤ 12 yr edu

Appendix 4: Components of the MoCa test recorded by the recruited speaker

Sounds/letters	A	B	C	D	E	F	J	K	L	M	N	O
Words	Face	Velvet	Red	Daisy	Church							
Sequences of numbers	2 1 8 5 4	7 4 2										
Sentences	I only know that John is the one to help today	The cat always hid under the couch when dogs were in the room										
Instructions	Hello this is a test sentence											

Appendix 5: Survey used in the experiment

Thank you! Now, we ask you to answer some brief questions. Please answer all questions below honestly and fully.

1. What is your age?
2. What is your gender? ○ Male ○ Female ○ Other ○ I'd rather not say
3. How many years of formal education have you had?
4. Please indicate what country and what region/state you currently live in.
5. Is your native language English? ○ Yes ○ No, my native language is:
6. Which languages can you speak apart from your native language?
Please also indicate at what age you started to learn that language/those languages.
7. Have you recently experienced any problems with reading? ○ Yes ○ No
8. Did you experience any reading problems during this experiment? ○ Yes ○ No
9. Have you recently experienced any problems with hearing? ○ Yes ○ No
10. Did you experience any hearing problems during this experiment? ○ Yes ○ No

Continue

Appendix 6: Feedback and comments page used in the experiment

Lastly, we give you the opportunity to give some feedback and commentary on this experiment.

1. Did you say true or false for this sentence: 'Some robins are birds'? Please explain why you chose that answer.

2. Did you say true or false for this sentence: 'Some ants are insects'? Please explain why you chose that answer.

3. Below this question you can see a table of all the words that were used in this experiment. Please write down the words that you were not familiar with during the experiment.

Fish	**Trees**	**Birds**	**Insects**	**Flowers**
Eels	Oaks	Sparrows	Flies	Roses
Carp	Beeches	Robins	Mosquitoes	Tulips
Herrings	Hazels	Magpies	Wasps	Chrysanthemum
Trout	Elms	Pigeons	Beetles	Daffodils
Sharks	Sycamores	Canaries	Ants	Daisies
Cod	Firs	Blackbirds	Bees	Lilies
Pike	Birches	Parrots	Worms	Carnations
Salmon	Willows	Crows	Ladybugs	Crocuses
Tuna	Poplars	Swallows	Roaches	Sunflowers
Goldfish	Pines	Seagulls	Moths	Orchids
Flounders	Chestnuts	Doves	Crickets	Hyacinths
	Cedar trees	Hummingbirds		Pansies
	Palm trees			Poppies

4. Do you have any other comments? This can be anything.

OK

Appendix 7: Model summary outputs

	Effects on Response to Underinformative Sentences		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	157.72	31.01 – 802.15	**<0.001**
High-load condition	1.48	0.39 – 5.53	0.564
Low-load condition	0.24	0.10 – 0.58	**0.001**
Age group old	11.25	2.29 – 55.31	**0.003**
High-load condition * Age group old	1.88	0.53 – 6.69	0.328
Low-load condition * Age group old	0.29	0.12 – 0.72	**0.008**
Random Effects			
σ^2	3.29		
τ_{00} Subject	10.18		
τ_{00} Sentence	0.59		
τ_{11} Subject.High-load condition	0.73		
τ_{11} Subject.Low-load condition	2.36		
τ_{11} Subject.Age group old	25.25		
τ_{11} Subject.High-load condition:Age group old	3.47		
τ_{11} Subject.Low-load condition:Age group old	2.11		
τ_{11} Sentence.High-load condition	0.02		
τ_{11} Sentence.Low-load condition	0.11		
τ_{11} Sentence.Age group old	0.20		
τ_{11} Sentence.High-load condition:Age group old	0.15		
τ_{11} Sentence.wmCondition2:AgeGroup1	0.18		
ρ_{01}	1.00		
	-0.63		
	0.63		
	0.45		
	-0.70		
	0.04		
	-0.90		
	0.87		
	0.80		
	-0.92		
N Subject	126		
N Sentence	30		
Observations	3537		
Marginal R^2 / Conditional R^2	0.702 / NA		