Older Adults' Comprehension of Distributive and Collective Quantification

(Bachelor's Project)

Maarten van der Velde, s2389894, m.a.van.der.velde.2@student.rug.nl
Internal supervisor: Dr. Jennifer Spenader[1]
External supervisor: Dr. Ken Drozd[2]

July 2015

Abstract

Sentences containing numerically quantified expressions (NQEs), such as 'Three people are holding two lamps', have multiple interpretations. Previous studies found that inserting 'each' causes young adults to strongly prefer a distributive reading, while five-year old children will readily accept both distributive and collective interpretations. This study investigated if adults over the age of 65 are similar to children in NQE interpretation due to age-related declines in working memory and inhibition. 24 older adults (mean age 74) participated in a truth-value judgement task with a 2 x 2 design. The factors were image type (distributive or collective) and sentence type (*3/2* or *3/each/2*). Additionally, working memory capacity and inhibition were measured. Unexpectedly, older adults' performance equalled or exceeded that of young adults. No effects of working memory capacity or inhibition were found.

1 Introduction

Research into language processing focuses largely on young adults or on children as they acquire language. However, with an increasingly ageing population knowledge about language processing in older adults has become an important topic of research.

We investigated whether ageing influences performance on interpretation of sentences like (1). Contrary to expectation, performance did not decline in adults aged 65 and over, but increased slightly compared to young adults. There was no negative impact of the older adults' lower working memory capacity and higher inhibitory deficits.

[1]Department of Artificial Intelligence, University of Groningen
[2]Center for Language and Cognition, University of Groningen

1.1 Interpretation of numerically quantified expressions

This study focused on the interpretation of sentences containing two *numerically quantified expressions* (NQEs). These sentences are ambiguous, allowing for multiple interpretations. Consider sentence (1).

(1) Three people are holding two lamps.

This sentence contains two NQEs: 'three people' and 'two lamps'. It is an example of a *3/2* sentence. The sentence can be given a *collective* reading (Figure 1a) or a *distributive* reading (Figure 1b). The collective reading of (1) says that there are three people who together are holding a single set of two lamps. The distributive reading says that there are three people who are each holding their own set of two lamps.

(a) Collective interpretation (b) Distributive interpretation

Figure 1: Two interpretations of sentence (1).

1.2 Disambiguation by adding 'each' or 'together'

The ambiguity of sentence (1) can be resolved by adding expressions such as 'each' and 'together'. These expressions force a distributive and a collective reading, respectively.

Consider sentence (2). It modifies sentence (1) by adding 'each'. The modified sentence is an example of a *3/each/2* sentence. In this sentence 'each' refers to the subject ('three people'), which takes scope over the object ('two lamps'). Because 'each' marks distributivity, the reader is forced to interpret the sentence distributively (as in Figure 1b).

(2) Three people are each holding two lamps.

'Together', on the other hand, marks collectivity. Consider sentences (3a) and (3b). The sentences differ on the position of 'together'. Nevertheless, in both cases 'together' refers to the subject ('three people'), which once again takes scope over the object ('two lamps'). The result is that only a collective reading is possible (as in Figure 1a).

(3) a. Three people together are holding two lamps.

b. Together three people are holding two lamps.

1.3 Age-related differences in NQE interpretation

Most previous studies have shown that young children are more lenient in the acceptance of certain collective and distributive interpretations than adults. In particular, there is evidence that young children are more likely than adults to accept distributive interpretations of *3/2* sentences and collective interpretations of *3/each/2* sentences.

This was illustrated by Drozd and Van der Lely (2014), who showed that five-year-old children accepted a distributive reading of an English *3/2* sentence such as (1) 93% of the time in a truth-value judgement task. In contrast, adults only accepted such an interpretation 53% of the time, a significant difference.

However, this difference was not present in an earlier study with a similar set-up by Musolino (2009), who found an acceptance rate of *3/2* distributive readings of around 80% by adults as well as five-year-old children. Drozd and Van der Lely (2014) suggested that both studies do show that adults prefer a non-distributive reading. Nevertheless the significant difference between children and adults is only present in Drozd and Van der Lely's (2014) study.

Musolino (2009) also studied *3/2* sentences in a collective context. These were fully accepted by the five-year-old children as well as the adults.

Syrett and Musolino (2013) performed a truth-value judgement task that included English sentences with 'each', a distributivity marker. As expected, adults fully accepted *3/each/2* sentences such as (2) when given a distributive reading. Five-year-old children did too. However, there was a significant difference when it came to a collective reading of *3/each/2* sentences: while adults only accepted this reading 31.9% of the time, five-year-old children did so 86.7% of the time. Therefore this is another factor combination for which there is a large difference in acceptance rate between adults and children.

Despite there being multiple studies of collective and distributive interpretation, there are still few explanations of the underlying phenomenon.

Pagliarini, Fiorin and Dotlačil (2012) argued that adults reject the distributive reading of definite plural noun phrases (without a distributive marker) like 'three boys' because of

a conversational implicature. They reasoned that such a phrase will not be interpreted distributively because adults expect the speaker to use 'each' if a distributive reading is intended. This is because the phrase 'each boy' is more specific and forces a distributive reading. The researchers suggested that very young children, who have not yet acquired understanding of 'each', will not apply this reasoning. They found evidence for this claim by showing that, as young Italian children acquired knowledge of 'ciascuno' ('each'), their acceptance of distributive readings of definite plural noun phrases (i.e., without 'each') decreased. Between the ages of four and seven, children still accepted such phrases fully. Thirteen-year-old children only accepted these phrases 72% of the time. Finally, the adult participants (mean age 21) really did not like these interpretations, only accepting them 50% of the time.

Drozd and Van der Lely (2014) followed similar reasoning, saying that when adults were asked to judge a *3/2* sentence in a distributive context, they tended to come up with a more specific sentence containing 'each' that they preferred. They rejected the more general *3/2* sentence. Both Musolino (2009) and Drozd and Van der Lely (2014) suggested that the difference in acceptance rate between adults and children may be due to the fact that children are less inclined to consider alternative sentences than adults. They proposed that children find it hard to revise an initial (incorrect) parse of an ambiguous sentence, unlike adults.

In summary, the studies described above have shown that children differ from adults in their abnormally high acceptance rates of distributive *3/2* readings and collective *3/each/2* readings.

Such differences in NQE interpretations may also be present when older adults are compared to young adults. Cognitive decline in older adults could cause them to have similar acceptance rates to young children.

1.4 Language and ageing

A popular theory to account for overall cognitive decline in older adults is that of general slowing. It says that the processing speed of individual operations declines with increasing age, leading to impaired cognitive functioning across a wide variety of tasks. The amount of slowing differs between processes, meaning that some are more affected than others (Salthouse, 1996).

Language, as a high-level cognitive function, may be expected to be subject to general slowing in older adults. However, age-related cognitive decline is not uniform across all areas of cognition. In fact, some functions, e.g. those pertaining to highly practised skills and familiar information, are quite well-preserved. Certain aspects of language, in particular language comprehension, fall in this domain. Adults' knowledge of words and their meanings does not decline as they get older. On the other hand, parts of language production, such as word finding and spelling, are very much affected by ageing (Burke & MacKay, 1997).

The non-uniformity of the decline of language functions has invoked a number of different explanations.

Resource theory, introduced by Kahneman (1973), says that all mental processes share a finite pool of (attentional) resources. These resources tend to be lower in older adults, leading to reduced performance on demanding cognitive tasks (Craik & Byrd, 1982). This would mean that language processes that require a lot of resources are more sensitive to age-related deficits than those that do not. NQE interpretation can be regarded as a high resource task as it involves assessing multiple readings of a sentence, which requires sufficient attentional control to analyse these readings, as well as the resources to maintain them. Therefore successful interpretation of NQEs may be negatively affected by resource deficits in older adults. The vagueness in defining 'resources' makes this hard to verify, however.

Similarly, a decline in general working memory capacity may influence some language processes more than others. Just and Carpenter (1992) proposed that working memory capacity and language comprehension are directly related. They argued that a larger capacity would increase the ability to maintain

multiple interpretations of an ambiguous sentence. Accordingly, Miyake, Just and Carpenter (1994) found that subjects with a regular to large working memory capacity were better able to maintain multiple interpretations of an unresolved lexical ambiguity than subjects with a small working memory capacity. Interpretation of NQEs is a similar process in that multiple interpretations of a sentence are analysed to resolve its ambiguity. This suggests that NQE interpretation may also be impacted by older adults' lower working memory capacity. Furthermore, Kemper and Kemptes (2002) found that older adults were less able to comprehend and produce complex syntax, as well as other complex semantic content. The reduction in performance was related to a decline in general working memory capacity.

However, this view was contested by Waters and Caplan (2001), who argued that on-line syntactic processing in sentence comprehension (i.e., as the stimulus is encountered) relies on a dedicated separate-sentence-interpretation resource (SSIR). They did not find a relation between efficiency in processing sentences of varying complexity and general working memory capacity. Additionally, unlike off-line processing (e.g. making a judgement after the stimulus has been presented), on-line processing efficiency also did not vary between age groups. This would suggest that NQE interpretation, being an on-line syntactic process, takes place in the SSIR and is unaffected by older adults' working memory decline.

Another view is that of inhibition deficit theory, which says that the strength of inhibitory processes declines with age. Lustig, Hasher and Zacks (2007) proposed that inhibitory deficits have a negative effect on the control of attention and working memory, causing a decline in processes that rely on analysing particular details while ignoring irrelevant information. Inhibitory deficits also reduce the ability to suppress strong initial responses to stimuli that may be incorrect. Both of these properties may be relevant to the process of NQE interpretation. For example, successful assessment of multiple interpretations of an ambiguous sentence containing NQEs may be impaired by an inability to select the relevant information and inhibit the rest. Also, when multiple interpretations of an NQE are entertained, but one is found more quickly than the others, inhibitory deficits may lead to the immediate acceptance of the first without really considering the other, 'slower' interpretations.

The differences between young children and adults described in section 1.3 may be caused by children's poor understanding of 'each' (Pagliarini et al., 2012).

However, the differences may also be rooted in cognitive function. Children may suffer more from processing difficulties when interpreting NQEs than young adults do. Given the cognitive decline in older adults, there may be parallels between young children and older adults in NQE interpretation.

Just like older adults, young children typically have a smaller working memory capacity than young adults (Gathercole, Pickering, Ambridge & Wearing, 2004). As was discussed in Section 1.4, sufficient working memory capacity was found to be crucial in maintaining multiple interpretations of an ambiguous sentence. This means that young children and older adults may display similar deficits in comprehension of NQEs.

Similarly, inhibition deficits may play an important role in selecting the correct interpretation for young children as well as older adults.

1.5 Research question

The aim of this study was to find if such parallels between young children and older adults' language comprehension exist. If they did, older adults would deviate from young adults' acceptance rate of ambiguous sentences in collective or distributive contexts in the same manner that young children do.

To that end, the following research question was posed:

Is there a difference between older adults and young adults in the acceptance rate of collective and distributive interpretations of ambiguous sentences containing numerically quantified expressions?

We proposed that older adults may show

preferences similar to children in their interpretation of such ambiguous sentences, because of similar limitations to working memory capacity and inhibitory processes.

Specifically, we expected very high acceptance rates for distributive readings of *3/2* sentences and collective readings of *3/each/2* sentences. Both of these combinations were not accepted well at all by young adults, but young children accepted them almost fully.

The research question was addressed by asking older adults to perform a truth-value judgement task (TVJT). In this task the participants decided for *3/2* and *3/each/2* sentences whether they matched images with a distributive or collective context.

To investigate the relation with working memory capacity, participants also performed a forward and backward digit span test. This test measures both the processing and storage capacity of working memory, making it correlate well with language comprehension (Daneman & Merikle, 1996).

Additionally, to investigate the effect of inhibition, a Stroop colour-word test was performed (Stroop, 1935).

2 Methods

2.1 Participants

24 native Dutch speakers (15 female, age range 65-89 years, mean age 74 years) participated. They did not receive any compensation. All participants gave informed consent before taking part in the experiment. All participants reported that Dutch was their first language. None of the participants reported having dementia, multiple sclerosis, or other brain diseases. All participants had normal colour vision and adequate hearing.

2.2 Design

The experiment had a 2 x 2 factorial within-subject design. The two manipulated factors were (1) image type (either distributive or collective) and (2) sentence type (either *3/2* or *3/each/2*).

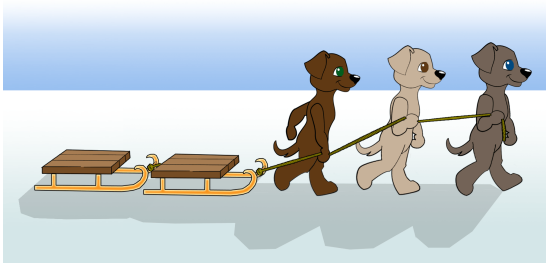

Figure 2: Example trial (collective *3/2*). Sentence: 'Drie honden trekken twee sleeën.' ('Three dogs are pulling two sledges.')

There were six trials for each of the four conditions, 24 in total. An additional 18 filler trials and six control trials were included. The experiment consisted of one block, containing all 48 trials presented in a random order. Four balanced lists were created.

The images and sentences in the experimental trials described 24 unique scenarios. Each scenario consisted of a subject, a verb and an object. The scenarios were balanced with regard to the object and verb. An example trial (collective image, *3/2* sentence) is shown in Figure 2.

The 18 filler trials were similar in design to the experimental trials, except for the sentence types. In these trials the Dutch word 'samen' ('together') was added to a *3/2* sentence in one of two places: at the start (as in sentence (3b)), or between the subject and the verb (as in sentence (3a)). As in the experimental trials, the images were either distributive or collective.

The six control trials served to determine whether a participant understood the task and was able to perform it successfully. These trials contained simple sentences consisting of a singular subject, a verb and a singular object. In three of the trials the sentence corresponded to the image. In the other three there was a clear mismatch, because either the subject or the object of the sentence did not correspond to the image.

A full list of the trials is included in Appendix A.

2.3 Materials

2.3.1 Stimuli and Apparatus

In each trial the participant was presented with an image. At the same time a spoken sentence was played. A button below the image could be clicked to replay the sentence. At the bottom of the screen were three buttons: *Yes* and *No* to select an answer, *Next* to confirm the answer and continue to the next trial.

The experiment was presented in a quiet room on a laptop computer in full screen. Sound was played through headphones or the laptop's speakers. Participants selected their answers using an external computer mouse or the computer's touch screen.

2.4 Procedure

2.4.1 Cognitive tests

Before the main experiment, participants completed a digit span test and a Stroop colour-word test.

The forwards and backwards digit span test was performed to investigate participants' working memory capacity (Turner & Ridsdale, 2004).

In addition, the Stroop colour-word test was used as a measure of executive function. The version used consisted of three cards: a word card, a colour card and a colour-word card. Each card contained 100 stimuli (Hammes, 1973).

2.4.2 Main task

The distributivity-collectivity truth-value judgement task took about 15 minutes. Participants were informed that they could take as much time as needed, and that they could take a break at any moment. Participants who were unable to operate the computer were assisted by the experimenter.

The experiment started with an instruction screen, informing the participant of the task and how to operate the interface. These instructions are shown in Appendix C.

The instructions were followed by a practice trial. The participant was allowed to ask questions about this trial to the experimenter. The audio volume was adjusted if necessary.

Next the participant was presented with a screen reminding them of the task instructions.

After this the actual experiment started, consisting of a single block of 48 trials. Participants were not allowed to ask questions about these trials.

Finally a screen was shown thanking the participant.

2.5 Analysis

2.5.1 Digit span test

The total number of correctly repeated sequences for both conditions was converted into a standard score based on Turner and Ridsdale (2004).

2.5.2 Stroop colour-word test

The Stroop test yielded six values per participant: the time needed to read each of the three cards, and the number of errors per card. A single score was generated by subtracting the time needed for the colour card from the time needed for the colour-word card. This scoring formula most accurately represents the interference factor measured by the Stroop task (Jensen, 1965).

2.5.3 Main task

Mean acceptance rates across participants for each of the conditions were calculated from the data. Data from participants who made more than one error on the six control trials were excluded from further analysis.

2.5.4 Generalised mixed effects model

A generalised mixed effects model (Bates, Maechler, Bolker & Walker, 2014) was made using the data of all three tests. A stepwise variable deletion procedure reduced the initial full interaction model to one that was parsimonious.

3 Results

3.1 Acceptance rates

Of the 24 participants tested, one was excluded from further analysis because of too many errors (2 out of 6) in the control trials.

The proportion of 'yes' answers across all trials was 51.9%. Therefore there was no bias towards one of the answers.

Figure 3 shows the mean acceptance rate across participants per condition.

3.1.1 Collective 3/2

Older adults accepted *3/2* sentences in combination with a collective image almost universally (94.9%). This result was entirely consistent with previous results from young adults as well as five-year old children (Musolino, 2009), confirming our expectations.

3.1.2 Collective 3/each/2

The mean acceptance rate of *3/each/2* sentences combined with a collective image was only 13.0% among older adults. Interestingly, this number is lower than the 31.9% found for young adults by Syrett and Musolino (2013), and much lower than the 86.7% found for five-year old children in the same study. This result contradicts the hypothesis, which predicted the older adults' acceptance rate to be similar to that of children. A possible explanation for the age-related difference may be rooted in conversational implicature, which is discussed in section 4.3.

3.1.3 Distributive 3/2

3/2 sentences combined with a distributive image had an unexpectedly low acceptance rate of 41.3% among the older adults participating in this study. In comparison, Drozd and Van der Lely (2014) found *distributive 3/2* acceptance rates of 53% for young adults and 93% for children. Furthermore, it is much lower than the approximately 80% Musolino (2009) found for both of these groups. According to our hypothesis the older adults' acceptance rate should have been around 80-90% as well. In fact, the acceptance rates in this condition followed the same pattern as with collective *3/each/2* readings: acceptance rate was inversely correlated with age. In both

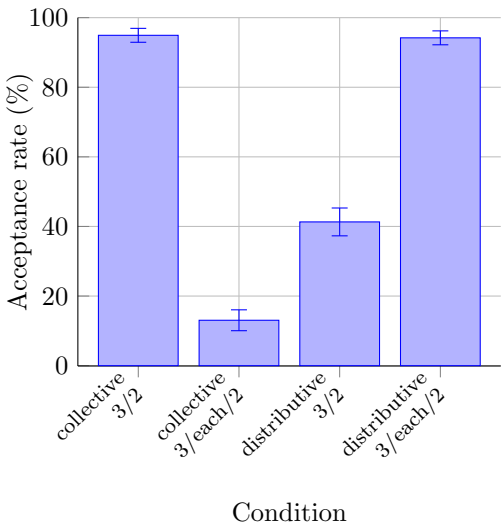

Figure 3: Mean acceptance rate across participants per condition. The error bars depict standard errors of the mean.

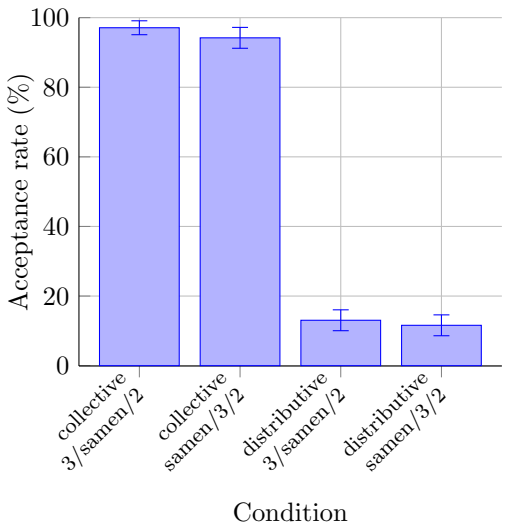

Figure 4: Mean acceptance rate across participants per condition of the filler trials. The error bars depict standard errors of the mean.

conditions, acceptance by children was near-universal, while young adults' acceptance rates were much lower and older adults accepted the interpretations even less frequently. Conversational implicature may again offer an explanation. This is examined in section 4.3.

3.1.4 Distributive 3/each/2

Finally, the combination of *3/each/2* sentences with distributive images had an acceptance rate of 94.2% among older adults. This finding was in accordance with our expectations. Syrett and Musolino (2013) found similar acceptance values, both for children and for young adults.

3.2 Filler trials

The mean acceptance rate across participants per condition of the filler trials is displayed in Figure 4 on the previous page. As expected, regardless of its position, sentences that included 'together' were completely accepted in a collective context (95.7%) and almost completely rejected in a distributive context (12.3%).

3.3 Cognitive tests

All of the 23 participants whose data was analysed also completed the forward and backward digit span test. The mean standard score on this test was 84.13 ($SE = 2.00$), which corresponds with the 14th percentile of the general population (Turner & Ridsdale, 2004).

In addition, 19 of the 23 participants did the Stroop colour-word test. The mean interference score across participants was 50.44 seconds ($SE = 4.17$ seconds). Van der Elst, Van Boxtel, Van Breukelen and Jolles (2006) found slightly higher mean interference scores of 60.92 seconds ($SE = 3.86$ seconds) and 58.54 seconds ($SE = 3.22$ seconds) for averagely educated 75-year old Dutch males and females, respectively. The same study found a mean interference score of 36.16 seconds ($SE = 1.84$ seconds) among averagely educated 25-year old Dutch males and 30.13 seconds ($SE = 1.18$ seconds) among averagely educated 25-year old Dutch females. This suggests that the participants had better inhibitory control than the general population of the same age, although it was still much lower than that of young adults.

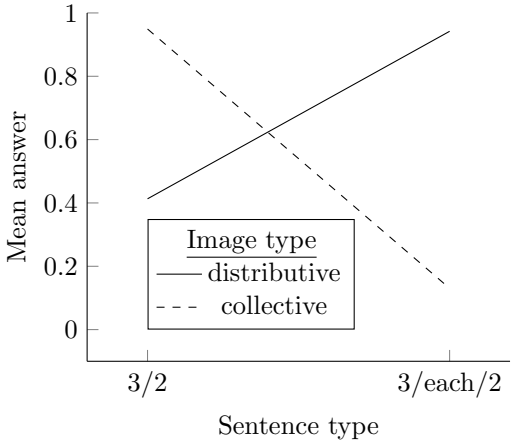

Figure 5: Interaction between sentence type and image type. An answer of 1 corresponds with 'yes', 0 with 'no'.

Condition I	Condition II	Significance
coll. 3/2	coll. 3/each/2	***
coll. 3/2	dist. 3/2	***
coll. 3/2	dist. 3/each/2	n.s.
coll. 3/each/2	dist. 3/2	**
coll. 3/each/2	dist. 3/each/2	***
dist. 3/2	dist. 3/each/2	***

Table 1: Significance of differences in means found in a Tukey's pairwise comparison of the conditions.

3.4 Generalised mixed effects model

The results were analysed using a generalised mixed effects model. Full details of this model are provided in Appendix B. The final model contained significant effects of sentence type ($p < 0.001$) and image type ($p < 0.001$). There was also a significant interaction between these predictors ($p < 0.001$), see Figure 5. To find the location of the interaction, a Tukey's pairwise comparison of means was performed. Table 1 shows the significance level of each difference found in this comparison.

The best model contained no significant ef-

fect of item, Stroop interference score, digit span score, age or gender. Nevertheless, the inclusion of both Stroop interference score and digit span score significantly improved the fit of the model ($\chi^2(2) = 55.076$, $p < 0.001$). For this reason these factors were included in the final model.

4 Discussion

This study investigated the interpretation of numerically quantified expressions by older adults.

Previous studies of young children and young adults revealed a relation with age when it came to acceptance of NQEs paired with a collective or distributive context in a truth-value judgement task. Children were much more accepting of certain interpretations that adults deemed wrong, in particular distributive interpretations of *3/2* sentences and collective interpretations of *3/each/2* sentences.

However, constructing a lifespan theory of NQE interpretation was outside the scope of these earlier studies, as they did not include older adults. This study aimed to fill that gap by investigating how adults over the age of 65 interpret NQEs.

Taking into account the decline with increasing age of certain aspects of cognitive performance, such as working memory capacity and executive control, the following research question was posed:

Is there a difference between older adults and young adults in the acceptance rate of collective and distributive interpretations of ambiguous sentences containing numerically quantified expressions?

We hypothesised that older adults' interpretations of NQEs would be more similar to those of children than to those of young adults. Age-related reduction of working memory capacity would impair older adults' ability to successfully comprehend and maintain multiple interpretations of an ambiguous sentence. This would lead to higher acceptance of interpretations that young adults reject, similar to children. Additionally, we expected older adults to have higher inhibitory deficits than young adults, resulting in more difficulty attending only to relevant information and suppressing incorrect knee-jerk reactions. This, too, would inflate acceptance rates of readings that young adults typically reject.

In answering the research question, this study found that there was indeed a difference between older adults and young adults. However, the results did not show the hypothesised similarity to children in the older adults' acceptance rates. In fact, in all conditions older adults' acceptance rates were equal to or lower than those of young adults. Section 4.3 offers a possible explanation for these results.

4.1 Working memory capacity

We hypothesised that older adults' relatively low working memory capacity would lead to acceptance rates more similar to those of children when judging sentences containing NQEs, since they would have more trouble maintaining multiple interpretations when compared to young adults with larger working memory capacities.

Unsurprisingly, the results from the forward and backward digit span test did show that the working memory capacity of the older adults in this study was much smaller than average. The mean score corresponded to the 14th percentile of the general population's scores.

Nevertheless, the older adults in this study did not appear to be held back by their lower working memory capacity. Their acceptance rates were similar to those of young adults, which suggests that working memory was not a bottleneck in the interpretation process. The mixed effects model confirmed that working memory capacity was not a significant predictor of acceptance rate.

Although previous studies suggested a correlation between working memory capacity and language comprehension (Just & Carpenter, 1992; Miyake et al., 1994; Kemper & Kemptes, 2002), in this study no such link was found.

On the other hand, the results could provide support for the existence of a dedicated separate-sentence-interpretation resource (SSIR), as proposed by Waters and Caplan

(2001). This SSIR is not related to working memory capacity and does not degrade with increasing age.

4.2 Inhibitory deficits

As inhibitory processes become less effective with increasing age, it becomes harder to control attention and working memory. We hypothesised that this increased interference could impair older adults' ability to carefully consider multiple interpretations of a sentence with NQEs. If this were the case, acceptance rates would deviate from those of young adults and be more like those of children. The results did not confirm this hypothesis.

As expected, the Stroop task scores showed that older adults had much higher inhibitory deficits than young adults. However, because the acceptance rates of both groups were similar, older adults did not appear to be hindered by these deficits. The mixed effects model also showed that Stroop task score was not a significant factor in predicting acceptance rate.

4.3 Conversational implicature

As Pagliarini et al. (2012) argued, adults expect a speaker to use 'each' when they wish to express distributivity. The absence of 'each' would cause the listener to doubt a distributive interpretation. The results of the older adults in this study are consistent with such a rule. When shown a distributive image and a *3/2* sentence, participants may have rejected it because they could think of a better fitting *3/each/2* sentence. This would explain why the acceptance of distributive *3/2* items was relatively low (41.3%).

Levinson (2000) postulated the Q principle, which states that as a speaker you must 'not provide a statement that is informationally weaker than your knowledge of the world allows' (2000, p. 76). This principle is violated whenever a speaker uses an informationally weaker *3/2* sentence instead of the stronger *3/each/2* form, while intending a distributive interpretation. Older adults appeared to be slightly more sensitive to violation of the Q principle than young adults and much more than children. Nevertheless, violation of this principle did not seem to completely rule out the possibility of a distributive interpretation, as even the older adults accepted nearly half of these items.

Just like young adults, older adults considered the presence of 'each' a marker for distributivity. As a result, *3/each/2* sentences with a collective image were only accepted 13.0% of the time. Once again older adults appeared to adhere to the rule more strictly than young adults: Syrett and Musolino (2013) reported that undergraduates had a collective *3/each/2* acceptance rate of 31.9%.

A problem of the conversational implicature account is that it does not explain why the presence of 'together', a collective marker, makes no difference to acceptance rate in the way that the presence of 'each' does. This is illustrated by the fact that collective readings of *3/2*, *together/3/2*, and *3/together/2* sentences were all accepted around 95% of the time by the older adults.

Furthermore, the presence or absence of 'each' cannot really be considered a scale, but is a division of pragmatic labour instead (Levinson, 2000, p. 137).

4.4 Directions for future research

Although there is a growing body of evidence for age-related differences in NQE interpretation, the underlying mechanisms are not yet well-understood. More research is needed to establish how robust the current findings are. Additionally, the role of cognitive resources like working memory and inhibitory control needs to be studied in more detail.

4.4.1 More participants

This study found no influence of working memory and inhibition decline on performance. However, it is too soon to completely rule out these factors. There were only 23 participants (and only 19 who did the Stroop inhibition test). One may still find an effect of cognitive decline on NQE interpretation with a larger sample.

4.4.2 Other age groups

The results of this study suggested that older adults are slightly more extreme in their rejection of *distributive 3/2* and *collective 3/each/2* items than young adults. As was discussed in section 4.3, this may be explained through the acquisition and practice of a conversational implicature rule. Compared to 20-year old young adults, adults around the age of 70 have had 50 years more experience in applying this rule. If the development of this skill is a constant process, one would expect middle-aged adults (between 40 and 50 years of age) to have acceptance rates in the middle between those of young adults and older adults. So far, middle-aged adults' interpretation of NQEs has not been studied.

Additionally, it may be beneficial to study NQE interpretation by adults over the age of 80 or 90 more extensively. Li and Baltes (2006) pointed out that performance on crystallised cognitive pragmatics (a practised language skill like NQE interpretation falls in this domain) peaks around 45 years of age and does not decline until after the age of 70. Singer, Verhaeghen, Ghisletta, Lindenberger and Baltes (2003) even found that aspects of verbal knowledge only started to decline at after age 90. Therefore there may be cognitive changes affecting NQE interpretation that only appear in very old age and were simply not present in our relatively young sample.

4.4.3 Cumulative context

Musolino (2009) found that both young adults and five-year old children fully accepted *3/2* sentences in a collective context. In a cumulative context[1], however, the adults' acceptance rate of *3/2* sentences was 78.1%, while the children's acceptance rate was only 23.4%. Drozd and Van der Lely (2014) also found a difference in acceptance rate of cumulative *3/2* items, although it was not significant: 68% for adults and 58% for children. Nevertheless, both studies appeared to show that a collective interpretation of a *3/2* sentence is preferred over a cumulative interpretation. This preference seems to be more pronounced in children than in young adults. Whether older adults also exhibit such a preference, and to what degree, is yet to be studied. Based on the results of this study, it is expected that older adults' acceptance rate of *3/2* cumulative items would be similar to or higher than that of young adults. This would follow the trend that with increasing age, the preference for collective over cumulative readings of *3/2* sentences decreases.

[1] For example, the cumulative reading of the *3/2* sentence 'Three people are holding two lamps' is that there are three people and two lamps, and every person is holding at least one lamp, and every lamp is being held by at least one person. The most extreme interpretation of this, in which every person is holding every lamp (and every lamp is being held by every person), is the collective reading.

4.4.4 Role of cognitive resources

Additional research is required to further investigate the role of working memory and inhibition when interpreting NQEs. One could assume that differences in NQE interpretation between young adults and children are (at least partly) due to age-related differences in working memory capacity and inhibition. Based on this assumption, young adults would be expected to have acceptance rates more similar to children when tested in a dual-task paradigm, where a secondary task is performed simultaneously to load the participant's cognitive resources. In this scenario, young adults would have to perform NQE interpretation with fewer available resources, making them more similar to children in terms of resources. If young adults' acceptance rates in the dual-task situation are more similar to those of children than in the single-task case, the cognitive resources occupied by the secondary task can be said to contribute to successful NQE interpretation. However, the findings from this study make such an outcome unlikely, as the older adults were seemingly not impaired by their cognitive decline.

4.5 Conclusion

Interpretation of ambiguous sentences containing NQEs was not found to be impaired by age-related cognitive deficits. Compared to young

adults, the older adults had a lower working memory capacity and higher inhibitory deficits. Nevertheless, they were still able to achieve young adult-like performance on the truth-value judgement task. In fact, older adults were even more critical than young adults of collective interpretations of *3/each/2* sentences and distributive interpretations of *3/2* sentences. Instead of the expected decline towards the level of children, older adults showed a slight improvement in interpretative skills when compared to young adults.

These findings counter the idea that performance on such language tasks is closely related to age or cognitive health. They underline the importance of carefully distinguishing between those functions that decline with age and those that do not. More research into the workings of the ageing brain is required to learn exactly what causes these differences in function degradation. The rapid growth of the population of healthy, older adults underlines the necessity of this research.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lme4` (R package version 1.1-7)

Burke, D. M. & MacKay, D. G. (1997). Memory, language, and ageing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1363), 1845–1856.

Craik, F. I. M. & Byrd, M. (1982). Aging and cognitive deficits. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (p. 191-211). New York: Springer US.

Daneman, M. & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422-433.

Drozd, K. & Van der Lely, H. K. (2014). *Children's comprehension of distributive universal quantification.* Unpublished.

Gathercole, S. E., Pickering, S. J., Ambridge, B. & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, *40*(2), 177.

Hammes, J. (1973). *De stroop kleur-woord test: Handleiding.* Amsterdam: Swets & Zeitlinger.

Jensen, A. R. (1965). Scoring the stroop test. *Acta Psychologica*, *24*, 398–408.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kemper, S. & Kemptes, K. A. (2002). Limitations on syntactic processing. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (p. 79-105). New York: Springer US.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* MIT Press.

Li, S.-C. & Baltes, P. B. (2006). Cognitive developmental research from lifespan perspectives: the challenge of integration. *Lifespan cognition: Mechanisms of change*, 344–363.

Lustig, C., Hasher, L. & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". *Inhibition in cognition*, *17*, 145–162.

Miyake, A., Just, M. A. & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of memory and language*, *33*(2), 175–202.

Musolino, J. (2009). The logical syntax of number words: Theory, acquisition and processing. *Cognition*, *111*(1), 24-45.

Pagliarini, E., Fiorin, G. & Dotlačil, J. (2012). The acquisition of distributivity in pluralities. In A. K. Biller, E. Y. Chung & A. E. Kimball (Eds.), *Proceedings of the 36th annual boston university conference on language development* (pp. 387–399).

Cascadilla Press.

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

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U. & Baltes, P. B. (2003). The fate of cognition in very old age: six-year longitudinal findings in the berlin aging study (base). *Psychology and aging*, *18*(2), 318.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Syrett, K. & Musolino, J. (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, *20*(4), 259-291.

Turner, M. & Ridsdale, J. (2004). *The digit memory test.* http://www.dyslexia-international.org/content/Informal tests/Digitspan.pdf.

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J. & Jolles, J. (2006). The stroop color-word test influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*(1), 62–79.

Waters, G. S. & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and aging*, *16*(1), 128.

Appendix A List of trials

A.1 Experimental trials

All sentences were used in two forms: *three subjects are verbing two objects*, and *three subjects are each verbing two objects*. Each form was combined with a distributive and collective image, resulting in four versions of each trial. The version of a trial a participant encountered depended on the list number. A complete overview of the subject-verb-object combinations used in the trials is given in Table 2.

A.2 Filler trials

The 18 filler trials are listed in Table 3. In 9 of these, 'together' was placed at the start of the sentence, while in the other 9 it was placed between the subject and the verb. 12 of the sentences were paired with a distributive image, 6 with a collective image.

Subject	**Verb**	**Object**
aliens	carry	ladders
aliens	hold	flags
aliens	pull	trailers
aliens	push	chairs
clowns	carry	mirrors
clowns	hold	presents
clowns	pull	toy cars
clowns	push	wardrobes
dogs	carry	tyres
dogs	hold	bones
dogs	pull	sledges
dogs	push	barrels
elephants	carry	trees
elephants	hold	branches
elephants	pull	boats
elephants	push	rocks
children	carry	suitcases
children	hold	pumpkins
children	pull	toy planes
children	push	hay bales
people	carry	paddles
people	hold	lamps
people	pull	book carts
people	push	boxes

Table 2: Subject-verb-object combinations used in the experimental trials.

A.3 Control trials

The sentences of the 6 control trials are listed in Table 4. These sentences had a singular subject and object. The first three trials had correctly matching images, while the images in the other three trials did not match the sentence due to an incorrect subject (4) or an incorrect object (5 and 6).

Subject	**Verb**	**Object**
aliens	carry	planks
aliens	pull	toy planes
aliens	push	boxes
clowns	hold	pumpkins
clowns	pull	trucks
clowns	push	hay bales
dogs	carry	loaves
dogs	hold	lamps
dogs	pull	book carts
elephants	carry	paddles
elephants	hold	flags
elephants	push	barrels
children	carry	ladders
children	pull	sledges
children	push	chairs
people	hold	branches
people	pull	locomotives
people	push	wardrobes

Table 3: Subject-verb-object combinations used in the filler trials.

Subject	**Verb**	**Object**
alien	carry	guitar
clowns	push	car
dogs	pull	carriage
woman	hold	cat
girl	carry	plank
man	hold	flag

Table 4: Subject-verb-object combinations used in the control trials.

Predictor	Estimate	SE	z value	p-value
(Intercept)	3.30827	0.57041	5.800	6.64e-09 ***
Sentence type (*3/each/2*)	-5.55826	0.69710	-7.973	1.54e-15 ***
Image type (*distributive*)	-3.54400	0.67858	-5.223	1.76e-07 ***
Stroop interference score	0.05102	0.29415	0.173	0.862
Digit span standard score	0.13529	0.27777	0.487	0.626
Sentence type (*3/each/2*) : Image type (*distributive*)	9.34344	0.89557	10.433	<2e-16 ***

Table 5: Fixed effects of the best-fitting generalised mixed effects model for acceptance rate with the factors image type and sentence type.
Formula:
Response ∼ Sentence type ∗ Image type + Stroop interference score + Digit span standard score + (1 + Image Type | Participant)

Appendix B Model

Table 5 contains the details of the best fitting generalised mixed effects model.

Appendix C Task instructions

The instructions (in Dutch) below were shown to the participant at the start of the distributivity-collectivity truth-value judgement task.

Taak:
U krijgt zometeen een aantal schermen te zien waarop telkens een afbeelding wordt getoond. Tegelijkertijd hoort u een zin.

Het is de bedoeling dat u telkens bepaalt of de zin bij de afbeelding past.

De afbeeldingen zijn simpele tekeningen van situaties die niet altijd waarheidsgetrouw zijn. Stelt u zich bij het beantwoorden een fantasiewereld voor waarin deze situaties zich afspelen. U dient dus niet te beoordelen of de situatie in de echte wereld zou voorkomen, maar alleen of de zin en de afbeelding bij elkaar passen.

Besturing:
Bij elke afbeelding hoort u de zin één keer. U kunt de zin herhalen door op de knop met de luidspreker te drukken. Dit mag u zo vaak doen als u wilt.

Vindt u dat de zin bij de afbeelding past, druk dan op '**Ja**'.
Vindt u dat de zin niet bij de afbeelding past, druk dan op '**Nee**'.

U kunt uw antwoord zo vaak als u wilt veranderen door alsnog op de knop met het gewenste antwoord te drukken.

Bent u zeker van uw antwoord, druk dan op '**Volgende**'. Deze knop legt uw antwoord vast en brengt u naar de volgende afbeelding.

Tijd:
U kunt bij het antwoorden zoveel tijd nemen als nodig is.

Het experiment duurt ongeveer 15 minuten. Het is op elk moment mogelijk om een korte pauze in te lassen. U kunt dan het scherm open laten staan.

Er volgt nu eerst een afbeelding om te oefenen. Uw antwoord telt nog niet mee. Druk op '**Oefenen**' om met de oefening te starten.