Acta Psychologica 229 (2022) 103690

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Individual differences in working memory and semantic fluency predict younger and older adults' multimodal recipient design in an interactive spatial task

Louise Schubotz [a], Aslı Özyürek [a,b,c], Judith Holler [a,c,*]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[b] Centre for Language Studies, Radboud University Nijmegen, the Netherlands
[c] Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Co-speech gesture
Cognitive aging
Recipient design
Common ground
Multimodal language

ABSTRACT

Aging appears to impair the ability to adapt speech and gestures based on knowledge shared with an addressee (*common ground-based recipient design*) in narrative settings. Here, we test whether this extends to spatial settings and is modulated by cognitive abilities. Younger and older adults gave instructions on how to assemble 3D-models from building blocks on six consecutive trials. We induced mutually shared knowledge by either showing speaker and addressee the model beforehand, or not. Additionally, shared knowledge accumulated across the trials. Younger and crucially also older adults provided recipient-designed utterances, indicated by a significant reduction in the number of words and of gestures when common ground was present. Additionally, we observed a reduction in semantic content and a shift in cross-modal distribution of information across trials. Rather than age, individual differences in verbal and visual working memory and semantic fluency predicted the extent of addressee-based adaptations. Thus, in spatial tasks, individual cognitive abilities modulate the interactive language use of both younger and older adults.

1. Introduction

Mutually shared knowledge between a speaker and an addressee (their *common ground*, Clark, 1996) affects how speakers speak and gesture in interaction with others. Previous research suggests that this addressee-based adaptation of utterances, or *recipient design* (Sacks et al., 1974), is modulated by normal human aging, such that older adults are less capable of engaging in successful recipient design than younger adults (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Saryazdi et al., 2019). Recent work employing a face-to-face, narrative setting suggests that this extends to the gestures accompanying speech (Schubotz et al., 2019). However, it remains unclear whether these behavioral differences in speech and gesture are connected to age-related differences in cognitive abilities and whether they also manifest in other communicative settings. Therefore, the aim of the present study was to investigate the extent of older and younger adults' verbal and gestural recipient design in a face-to-face interactive, spatial task, and to determine whether the communicative behavior in this context is modulated by age-related differences in cognitive abilities.

1.1. Age-related differences in verbal and gestural recipient design

Previous research suggests that there are systematic differences in how younger and older adults adapt their speech and their co-speech gestures based on knowledge shared with an addressee. Younger and older adults' addressee-related adaptations in the spoken modality (verbal recipient design) have been investigated using referential communication tasks, in which common ground builds up gradually as a result of repeatedly referring to the same set of referents over the course of several trials (*incremental* common ground, e.g. Clark & Wilkes-Gibbs, 1986; Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Younger adults are generally found to interact increasingly more efficiently, indicated by shorter utterances and fewer dialogue turns on later compared to earlier trials, as common ground accumulates. Although older adults' interactions follow the same general pattern of reduction, their interactions are characterized by longer utterances, more dialogue turns, and/or more errors relative to younger adults (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Additionally, unlike younger adults, older adults failed to produce

* Corresponding author at: Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands.
E-mail address: Judith.holler@mpi.nl (J. Holler).

https://doi.org/10.1016/j.actpsy.2022.103690
Received 18 August 2020; Received in revised form 20 July 2022; Accepted 21 July 2022
Available online 9 August 2022

appropriate common ground-based utterances in a subsequent task which involved familiar and new addressees (Horton & Spieler, 2007). However, evidence to the contrary was obtained by Yoon and Stine-Morrow (2019) in a live, interactive referential task.

Younger and older adults' addressee-related adaptations in the gestures accompanying their speech (gestural recipient design) have so far only been compared in one recent study (Schubotz et al., 2019). As interactive, face-to-face language use comprises of speech *and* co-speech gestures, taking this additional visual modality into account is crucial. Co-speech gestures are meaningful hand movements that accompany speech and contribute to the meaning of utterances in important ways via their semantic or pragmatic content (Kendon, 2004; McNeill, 1992). Schubotz et al. (2019) investigated younger and older adults' verbal and gestural adaptation to common ground, by asking younger and older participants to narrate short comic stories to a same-aged addressee. Mutually shared knowledge was induced by showing both participants part of the story at the start of each trial, while only the designated narrator saw the full story subsequently (a form of *personal* common ground, see Clark, 1996). Only younger, but not older adults, provided longer and more informative narrations and gestured at a higher rate when relating unknown as opposed to mutually shared story content. Older adults showed no evidence of common ground-based recipient design either in their speech or in their gestures, and even produced fewer rather than more gestures in relation to speech when relating novel story content in this task.

1.2. The role of cognitive abilities in recipient design

Such changes in communicative behavior, which may be taken to reflect the absence of appropriate recipient design, or the failure to take the addressee's knowledge state into account, have previously been speculated to be caused by cognitive aging, such as age-related deficits in working memory (WM) or executive control (e.g., Healey & Grossman, 2016; Horton & Spieler, 2007; Hupet et al., 1993; Long et al., 2018; Wardlow, 2013). WM may be involved in the ability to establish an addressee's perspective and to incorporate this perspective during online language processing, while executive control may be involved in the ability to inhibit one's own perspective in favor of the addressee's (e.g., Brennan et al., 2010; Keysar et al., 1998; see also Brown-Schmidt, 2009, for the role of executive function in perspective-taking during language comprehension).

Yet, Schubotz et al. (2019) could not establish a relationship between measures of verbal WM or executive control and older adults' lack of verbal and gestural recipient design. Although it is possible that the measures employed did not capture the abilities involved in the task at hand, they also considered the possibility that, beyond changes in cognitive abilities, differences in communicative goals may have determined how older adults design their utterances for others (e.g., Adams et al., 2002; Horton & Spieler, 2007; Underwood, 2010; see also Long et al., 2020, for an account of how differences in communicative strategies affect older adults' language use). While the younger participants in Schubotz et al. (2019) presumably focused mainly on information transfer, i.e. providing the addressee with information she did not yet have, older adults may have interpreted it as a task where it mattered to be a 'good story teller', in the sense of providing an easy-to-follow narrative and being clear and exhaustive in terms of story events. That is, aspects like the wish to narrate a nice and complete story may have overruled common ground-based adaptations of speech and gesture.

1.3. Spatial vs. narrative task demands during multimodal utterance design

One aspect which likely affects speakers' multimodal language use is the type of communicative task they wish to accomplish, e.g., whether this task is predominantly narrative (such as a story-telling task) or spatial (such as providing spatial descriptions or instructions). Associated task demands may not only modulate the use of the different modalities during utterance production, but also the extent to which WM and other cognitive resources are taxed. For example, observations of younger (e.g. Alibali, 2005) and older speakers (Feyereisen & Havard, 1999) show that people gesture more frequently when they talk about spatial topics, including visual and motor imagery, as opposed to verbal, abstract ones. This suggests that in spatial tasks, information is distributed differently across the two modalities. Gestures carry relatively more communicative weight and might therefore also be more relevant for successful recipient design. It is currently unclear how this distribution of information is organized, particularly for older adults, and how it is affected by pragmatic factors, such as common ground.

Furthermore, in spite of this higher gesture frequency observed during spatial descriptions, previous research involving visual and motor imagery in a monolog setting (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou et al., 2015) found older adults to produce relatively fewer depictive gestures than younger adults overall. Schubotz et al. (2019), on the other hand, found older adults to gesture as frequently as younger adults in the narrative task. It remains to be seen whether, given an interactive, face-to-face setting, older adults' gesture frequency is comparable to that of younger adults, also in a spatial task.

In addition to these direct effects on multimodal utterance production, visuo-spatial tasks may also differ from narrative tasks in terms of the involvement of cognitive abilities. Visuo-spatial tasks presumably rely more strongly on visuo-spatial cognitive abilities. While it appears that certain abilities such as visuo-spatial perception or mental imagery maintenance undergo only minor age-related decline, spatial WM is more strongly affected by age-related changes (for a review see Klencklen et al., 2012). Whether potential age-related differences in spatial cognition affect the use of spatial language remains unclear (see Markostamou et al., 2015).

Finally, previous research on younger adults suggests that gesticulation allows speakers to "off-load" information onto visual space, thereby freeing up cognitive resources more generally (e.g., Goldin-Meadow et al., 2001; Cook et al., 2012). This suggests that the potentially higher gesture rates associated with a spatial task, relative to a narrative task, might allow older adults to engage more cognitive resources towards recipient design.

In summary, it remains an open question how older adults' interactive use of speech and co-speech gestures in a visuo-spatial task compares to that of younger adults, and whether and to what extent communicative behavior, including multimodal recipient design, is modulated by cognitive abilities.

1.4. The present study

In the present study, we employed an interactive, spatial task in order to investigate whether the previously observed age-related differences in recipient design, spoken as well as gestured, extend to the spatial domain, and whether potential age-related differences in behavior can be attributed to differences in cognitive abilities.

In order to address these issues, we designed an interactive task in which a primary participant (the *speaker*) assembled little wooden castles from a set of building blocks and subsequently instructed a secondary participant (the *addressee*) on how to assemble the same castles. Mutually shared knowledge between speaker and addressee was manipulated per trial, by either showing both participants a picture of the to-be constructed model shortly at the beginning (common ground [CG]), or not (no common ground [no-CG]). We thus induced a form of personal common ground (Clark, 1996), in which the mutually shared knowledge existed from the outset of the interaction. Additionally, as speaker and addressee interacted over six consecutive trials, and the speaker referred to the same entities repeatedly, we also expected common ground relating to the individual building blocks and the steps of how to assemble them to build up incrementally (incremental common ground, Clark, 1996).

Apart from this additional possibility for incremental common ground to develop in the course of the experiment, the task employed here differed from the narrative task employed in Schubotz et al. (2019) in several other ways. First of all, as argued above, a visuo-spatial task differs substantially from a narrative task, both in terms of how information can be distributed across the two modalities, and in terms of the cognitive functions involved in the task. Furthermore, in the present task, the goal (i.e., give addressee instructions on how to assemble the model) allowed less room for individual interpretation and therefore minimized the likelihood that potential age-related differences in behavior could be attributed to age-related differences in task interpretation or communicative goals.

For our manipulation of personal common ground, we expected younger adults to produce fewer words and fewer gestures and to convey less information in the two modalities when providing instructions for previously seen models as compared to unseen models (Galati & Brennan, 2010, 2014; Holler & Wilkin, 2009; Schubotz et al., 2019). For the effects of incremental common ground, we expected the repeated references to the same entities and assembly steps over the course of the experiment to result in increasingly shorter utterances (e.g. Clark & Wilkes-Gibbs, 1986; Fussel & Krauss, 1992) and fewer gestures (Galati & Brennan, 2014), in concordance with a reduction in information content in both modalities. We made no directional predictions relating to gesture rate (i.e., number of gestures per 100 words), since this relation has been differently affected in previous studies (see Holler & Bavelas, 2017).

Due to the previously observed absence of verbal and gestural adaptations to personal common ground in older adults (Schubotz et al., 2019; also Horton & Spieler, 2007), we expected older adults to be less adaptive than younger adults to mutually shared knowledge induced on individual trials, not only in their speech but also in the way they draw on gesture when designing utterances for knowing vs. unknowing recipients. However, due to differences in task demands/design, particularly the spatial nature of the present task and the way in which the task goal was formulated, these age-related effects may be less pronounced than those obtained in Schubotz et al. (2019). In terms of verbal and gestural adaptations to incremental common ground, we expected older adults to show an overall pattern of reduction similar to that of younger adults, although this may be less pronounced than in younger adults (e.g., Hupet et al., 1993).

We additionally assessed how information was distributed across the two modalities, i.e. whether information was expressed uniquely in speech, uniquely in gesture, or whether it was expressed in both modalities. This provides an indication of the relative communicative weight that gestures carry and can additionally been seen as an indicator of recipient design: encoding the same piece of information twice, in both modalities, is arguably more informative than encoding information in only a single modality (see also de Ruiter et al., 2012). Therefore, we expected that younger participants would encode more information in both modalities, for unknowing as opposed to knowing addressees. Similarly, we expected younger adults to encode increasingly fewer pieces of information in both modalities, across the experiment, as common ground incrementally accumulates. Again, these effects may be smaller or absent in the older adults.

In order to test for the role of cognitive abilities in speech and co-speech gesture use and their adaptation to common ground, we assessed speakers' verbal and visual WM as well as executive control and semantic fluency. As summarized above, verbal WM and executive function have previously been related to verbal recipient design, as well as visual perspective taking (e.g. Hupet et al., 1993; Long et al., 2018; Wardlow, 2013). Our expectation was that higher verbal WM and executive function might be associated with more pronounced verbal and/or gestural recipient design and that potential age-related differences in verbal and gestural behavior may be attributable to age-related differences in these cognitive functions.

Furthermore, verbal and visual WM have previously been related to gesticulation in general (e.g. Chu et al., 2014 for visual WM; Gillespie et al., 2014 for verbal WM), such that lower cognitive abilities lead to higher gesture frequencies (for the cognitively beneficial effects of gesticulation see also e.g. Goldin-Meadow et al., 2001; Cook et al., 2012). Similarly, lower semantic fluency is an indicator of word finding difficulties, which in turn may be associated with an increase in gesticulation (Rauscher et al., 1996). We included these measures in order to be able to control for the possibility that potential age-related differences in the interactive use of co-speech gestures are attributable to age-related differences in these cognitive functions.

2. Method

2.1. Participants

The same participants as in Schubotz et al. (2019) participated in the present experiment: thirty-two younger adults (16 women) between 21 and 30 years old (M_{age} = 24.31, *SD* = 2.91) and 32 older adults (16 women) between 64 and 73 years old (M_{age} = 67.69, *SD* = 2.43). All participants were native Dutch speakers with self-reported normal or corrected-to-normal vision and hearing and no known history of neurological impairment. Each participant was allocated to a same-age and same-sex pairing. The role of speaker or addressee was randomly assigned and kept constant across the entire experiment. Only the speaker data was analyzed here. All participants in the role of speaker had minimally secondary school education, except for one older participant who only had primary school education. Participants were recruited from the participant pool of the Max Planck Institute for Psycholinguistics and received between € 8 and € 16 for their participation, depending on the duration of the session. The experiment was approved by the Ethics Commission for Behavioural Research from Radboud University Nijmegen. Signed consent was acquired from all participants.

2.2. Materials

We created six black-and-white line drawings of simple castle-like buildings (for examples see Fig. 1). Each castle could be assembled from seven wooden building blocks, all of the same color: two cubes, two rectangular prisms, two triangular shapes (right triangles), and one arc-shaped block, ranging in size from 4 × 4 × 4 cm (cubes) to 4 × 4 × 12 cm (arc). The buildings were constructed such that the two triangular shapes always formed the top of the building, the position of the remaining building blocks varied. All models were fully symmetrical. We intentionally kept the models simple in order to ensure that older

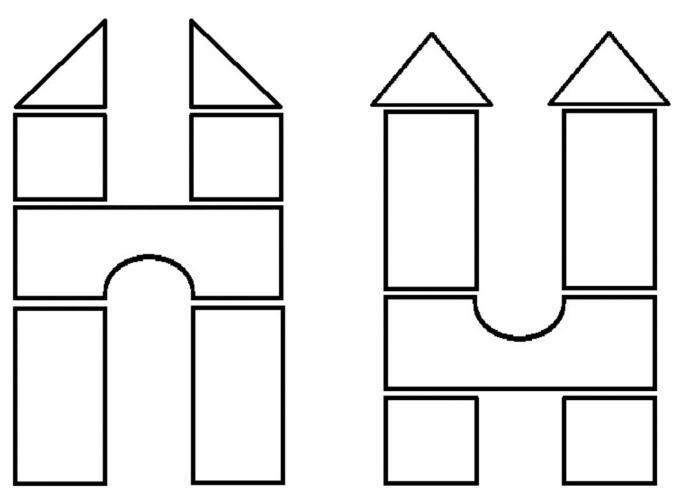

Fig. 1. Two of the six stimuli used in the experiment.

adults would be able to memorize them correctly in spite of potential age-related memory deficits.

Four experimental lists determined the order in which the different models were presented. Initially, we created two orders of presentation for the six models, one being the reverse of the other. Counterbalancing the order of common ground presentation across lists resulted in four experimental lists. Each list was tested eight times, distributed evenly across age groups and sexes.

2.3. Procedure

Upon arrival, the speaker and the addressee were asked to sit in designated chairs at a table at 90° from each other. Two video cameras were set up on tripods at a small distance from the table, one capturing a frontal view of the speaker, the other one positioned such that it captured both speaker and addressee (see Fig. 2 for stills from the two cameras). Sound was recorded with an additional microphone suspended from the ceiling over the table and connected to the speaker camera.

All participants completed one practice trial and six experimental trials. At the beginning of the practice trial and of half of the experimental trials (the CG trials), both participants were presented with a line drawing of a model and instructed to look at it carefully for 5 s without talking, with the aim to experimentally induce common ground about the composition of the model. Subsequently, the drawing was removed and a screen was put up on the table between speaker and addressee. The speaker then received the drawing and the seven building blocks and assembled the model according to the drawing. Once the speaker indicated that he/she was done, the experimenter checked the model for accuracy and then took the blocks and the drawing away. The screen was taken off the table and the speaker described to the addressee how to assemble the model, without using the building blocks. Addressees were instructed to listen to the descriptions and ask all clarification questions at the end. Once speaker and addressee had discussed potential questions, the screen was put back up and the addressee received the building blocks in order to assemble the model according to the speaker's instructions. Addressees built the model behind the screen in order to avoid any engagement of the speaker during this process. Additionally, the experimenter took away the construction built by the addressee before removing the screen, and feedback on the addressee's performance was given only at the end of the entire task. This was done in order to avoid any adaptation of the speaker's instructions based on the addressee's performance.

For the other half of the experimental trials (the no-CG trials), the procedure was identical, except for the first step: participants did not see the model picture beforehand, rather, the screen was put up at the beginning of the trial and the speaker received the model picture and the building blocks immediately. Depending on the pair, the task took about 20 to 30 min. After the experimental tasks were completed, the addressee was allowed to leave, while the speaker performed the cognitive tests.

2.4. Transcription and coding

2.4.1. Speech coding

All recordings from the two cameras were synchronized and subsequently segmented into trials. Transcription of speech and annotation of gestures was conducted in ELAN (Version 4.9.4, 2016; Wittenburg et al., 2006). For all segments, the speaker's initial instruction, i.e. the first complete instruction on how to assemble the model without potential subsequent repetitions, was identified. All analyses reported here are based on these initial instructions only, discarding repetitions or clarifications elicited by the addressee following the initial instruction, as the focus of our study was the effect of our experimental manipulations on the speakers' behavior rather than the impact of speaker-addressee interaction (for a similar argument see Horton & Gerrig, 2005). Speech from the speaker was transcribed verbatim, including disfluencies such as filled pauses and word fragments. However, disfluencies were excluded from the word counts presented in the results section, as we were mainly interested in speech content and did not want potential age-related differences in the number of disfluencies to influence the word count (e.g. Mortensen et al., 2006). For this reason, we also distinguished between "narrative speech" belonging to the instruction proper (i.e. relating to the model building itself) and "non-narrative speech", such as statements about the task or comments relating to the speaker or the addressee (for the basis of this distinction see McNeill, 1992). Only speech belonging to the instruction proper entered the word count.

2.4.2. Gesture coding

For the gesture coding, we first identified all co-speech gestures produced by the speaker during the instruction proper, disregarding non-gesture movements as well as gestures accompanying non-narrative speech. Our unit of analyses was the gestural *stroke*, i.e. the most meaningful part of the gesture determined according to criteria established in previous co-speech gesture research (Kendon, 2004; Kita et al., 1998; McNeill, 1992). We then categorized these strokes as *representational* and *non-representational* gestures (see Alibali et al., 2001). For our purposes, representational gestures include *iconic* gestures, which iconically depict shape or size of concrete referents or represent physical movements or actions; *metaphoric* gestures, which resemble iconic gestures but relate to speech in a metaphorical manner (e.g. a rotating movement of the hand to indicate the passing of time); and pointing gestures or *deictics*, i.e. finger or whole-hand points to a specific location in real or imaginary space, e.g. that of a building block (McNeill, 1992).

All other gestures were considered non-representational and include what are frequently called *beat* gestures, i.e. biphasic movements of the hand, for example to add emphasis, as well as *pragmatic* gestures (Kendon, 2004), i.e. gestures which have pragmatic functions, for example to convey information about how an utterance should be interpreted, or relating to the interaction with the addressee (Bavelas et al., 1992; Bavelas et al., 1995).

A second coder blind to the experimental hypotheses coded 10 % of

Fig. 2. Example of the lateral (left panel) and frontal (right panel) views of the speaker in the experimental set-up. In this frame, the speaker indicates the size and position of the two small cubes.

the trials randomly selected from across all participants for stroke identification, and another 10 % of the trials for gesture categorization. Inter-rater agreement on stroke identification was 90.99 %. Inter-rater agreement on gesture categorization was 96.43 %, Cohen's Kappa = 0.86.

2.4.3. *Representational gesture frequency and gesture rate (gestures per 100 words)*

As we were mainly interested in the semantic content of the descriptions and the accompanying gestures, in our analyses we focus exclusively on representational gestures (i.e., iconic, metaphoric, and deictic gestures). In addition to reporting the raw representational gesture frequency, we computed a gesture rate per 100 words (see above for criteria on word count) by dividing the number of gestures by the number of words a given participant produced for each trial and multiplied this by 100. This gesture rate normalizes for differences in instruction length (e.g. Alibali et al., 2001).

2.4.4. *Information content coding*

In order to assess whether age and common ground (both personal and incremental) affected the information content of the speakers' utterances, we additionally coded for semantic features expressed in speech and in gesture. Per block, several pieces of information could theoretically be encoded: extrinsic features like the block's location and its orientation, and intrinsic features like its shape and its size. The actual scoring of individual features depended on the modality (see below, see also Holler & Wilkin, 2009, for a similar approach to scoring semantic features in speech and co-speech gesture).

2.4.4.1. Coding of semantic features encoded in speech. For speech, we scored whether the verbal description contained information with respect to three categories, namely each block's *location* (e.g., “at the bottom”, “on top”), its *orientation* (e.g., “upside down”, “vertically”), and the intrinsic features *shape and size* (e.g., “triangle”, “square”, “long”, “short”). Note that more metaphorical descriptions like “bridge” or “roof” were not counted as conveying orientation or shape information, since these terms refer to objects that may take a variety of shapes and may therefore elicit different visual imagery in different people. For each feature, we scored “1” if the information was present (the maximum score per feature was always “1”, even if the information was repeated or rephrased) or “0” if the information was absent. For the small cubes (see Fig. 1), encoding its orientation was not possible, yielding a maximum score of eleven features per description (four blocks à three features, minus one). A second coder blind to the experimental hypotheses recoded 10 % of the trials. Inter-rater agreement on scoring of location and of orientation in speech was 100 % each, inter-rater agreement on scoring of shape/size in speech was 97.5 %, Cohen's Kappa = 0.92.

2.4.4.2. Coding of semantic features encoded in gesture. For gestures, we scored whether manual movements contained information with respect to the same semantic aspects but used just two categories, namely a block's *location* (e.g., pointing to a certain point in space, performing the gesture in the appropriate area in space; gestures had to be spatially coherent with respect to the actual model and with respect to each other) and its *orientation*, *shape*, or *size* (e.g., moving two fingers up and down to indicate a block's vertical orientation, tracing a triangle shape, using two fingers to indicate the size of a block). Unlike for speech, for gesture we collapsed orientation and shape/size, because gestures consistently expressed several aspects at the same time due to their holistic nature, making it difficult to score these aspects separately (e.g., tracing an arc indicates the shape, the size, and the orientation of the arc-shaped block all at the same time). As for speech, we scored “1” if the information was present (“1” was the maximal score, even if several gestures were used to convey different aspects of one feature, e.g. a block's shape) and “0” if the information was absent. Additionally, in the coding of location, we introduced half a score (“0.5”). This was used if the gesture itself encoded the correct location information, but could not be related to a previous gesture, e.g. because there was no previous gesture or because the previous gesture was not performed in the correct location. By introducing this penalty, we aimed to account for whether the descriptions were spatially coherent or not. The maximum score for gesture information was eight (four blocks à two features). The same coder who coded the information in speech also recoded the same 10 % of the trials for information content in gesture. Inter-rater agreement on location coding was 92.50 %, Cohen's Kappa = 0.87, inter-rater agreement on orientation/shape/size coding was 98.75 %, Cohen's Kappa = 0.97.

2.4.5. *Number of semantic features encoded in speech and in gestures*

For each description, we computed the sum of semantic features encoded in speech and the sum of semantic features encoded in gesture. These provide an index of the total information that a speaker provided in each modality for each description, independently of what was represented in the respective other modality. In the Supplementary materials, Supplementary materials Section A, we additionally present an analysis of the normalized counts, i.e. the sums of semantic features encoded in speech and in gesture divided by the number of words and gestures respectively, which provides a measure of “information density” and an index of how efficiently the two modalities are used.

2.4.6. *Distribution of information across speech and gestures*

Finally, we also computed how many features were expressed in a single modality, i.e. only in speech or only in gestures, and how many features were expressed in both modalities, e.g. by referring to the triangles in speech while at the same time tracing their shape with the fingers. Based on these counts we computed the percentages of information encoded uniquely in speech, uniquely in gesture, or in both modalities, which provides an index of how information is distributed across the two modalities.

Table 1
Mean scores (and *SD*) per age group on cognitive tests, plus statistical comparisons (independent *t*-tests and Mann-Whitney tests where appropriate).

	Younger	Older	Test statistic
Verbal WM (Operation Span Task)	44.06 (*8.39*)	34.73 (*8.92*)	$t(29) = 2.99$**
Semantic fluency (Animal Naming Test)	31.50 (*9.40*)	27.75 (*5.99*)	$t(30) = 1.35$
Inhibitory control (Trail Making Test, TMT)[a,b]	14.5 (*27*)	21.5 (*62*)	$W = 65.5$*
Visuo-spatial WM (Visual Patterns Test, VPT)[a]	13 (*4*)	10 (*8*)	$W = 187.5$***
Visuo-sequential WM (Corsi Block Task, CBT)[a]	54 (*42*)	37.5 (*34*)	$W = 215$***

*** $p < .001$.
** $p < .01$.
* $p < .05$.
[a] Owing to the non-normality of the data, the figures represent Median (and *Range*).
[b] Note that smaller numbers indicate better performance on this task.

2.5. Cognitive measures

Participants performed the Operation Span Task (Ospan) as a measure of verbal WM, the animal naming task to assess semantic fluency, the Trail Making Test (TMT) as a measure of executive function, the Visual Patterns Test (VPT) as a measure of visuo-spatial WM, and the Corsi Block Task (CBT) as a measure of visuo-sequential WM. Detailed descriptions of these cognitive tasks, how they were administered, and how the scores were computed can be found in the Supplementary materials, Supplementary materials Section B. The summary of test scores provided in Table 1 indicates that younger adults outperformed older adults on all measures, except for the semantic fluency test.

2.6. Statistical methods

To investigate the influence of age, personal common ground (CG vs. no-CG trials), and incremental common ground (operationalized as trial number), as well as their interaction effects on the main speech- and gesture-based measures (word and gesture count, gesture rate, information contained in speech, gesture, and speech and gesture combined), we fitted linear mixed-effect models in R version 3.2.1 (R Development Core Team, 2015), using the package lme4 (Bates et al., 2017). We used likelihood ratio tests for model comparisons, eliminating all non-significant interactions in the model comparison process. For each dependent measure, we only report the estimates, *SE*s, *t*-values and *p*-values for the main experimental predictors, as well as other significant predictors and interactions (if applicable). All the models reported contain random intercepts for participants and items, but no by-participant random slopes for the common ground manipulation or repeated interaction as this led to perfect correlations of random factors throughout. Reported *p*-values were obtained via the package lmerTest (Kuznetsova et al., 2016).

To investigate the influence of cognitive abilities on our main dependent measures, and to test whether potential age-related differences in verbal and gestural behavior could be attributed to age-related differences in cognitive abilities, we applied the same basic procedure as described above, creating separate models for each cognitive predictor. As the analyses were exploratory, we performed a backwards-model-stripping procedure, starting out with a full model including the z-scored cognitive predictor of interest, age, the common ground manipulation, and trial number, as well as all their interaction terms, eliminating non-significant interactions and predictors based on likelihood ratio tests. See Section C of the Supplementary materials for details of the models.

Post-hoc Bayesian statistics: In addition to the linear mixed effects models, we used independent Bayesian *t*-tests to compare the two age groups (old/young) to establish Bayesian *t*-tests were calculated using the BayesFactor package [ttestBF function for two independent groups, version 0.9.2+, with default settings], within the R environment (R Core Team, 2021) where *p*-values indicated the group difference not to be significant to further evaluate the evidence against the null hypothesis.

3. Analyses and results

3.1. Descriptive statistics

Out of the 96 instructions participants gave, six contained an error. In all cases, participants confused the position of the cubes with that of the rectangles. Exactly half of the errors occurred in descriptions by older adults, suggesting that overall, younger and older adults were able to memorize the castles equally well.

Younger adults produced a total of 692 gestures, out of which 618 were iconic gestures (89.31 %), 3 metaphoric gestures (0.43 %), 70 abstract deictic gestures (10.12 %), 1 concrete deictic gesture (0.14 %), and 61 non-representational gestures (8.82 %). Older adults produced a total of 722 gestures, out of which 651 were iconic gestures (90.17 %), 1 metaphoric gesture (0.14 %), 67 abstract deictic gestures (9.28 %), 3 concrete deictic gestures (0.42 %), and 19 non-representational gestures (2.63 %).

Table 2
Means and SDs for dependent measures by age group and personal common ground condition.

	Younger		Older	
	CG	No-CG	CG	No-CG
Number of words	42.4 (18.29)	50.71 (23.14)	39.79 (21.19)	48.94 (32.59)
Number of gestures	6.63 (3.87)	7.79 (5.69)	6.71 (5.21)	8.33 (8.54)
Gestures/100 words	17.24 (11.00)	15.04 (7.56)	18.69 (13.28)	19.14 (12.05)
Speech info total	7.29 (1.75)	7.54 (1.71)	6.44 (2.06)	6.6 (2.33)
Gesture info total	4.59 (2.54)	5.04 (2.67)	4.49 (3.16)	4.71 (3.06)
% Multimodal info	38.98 (22.74)	42.01 (24.95)	35.4 (25.3)	36.51 (26.23)
% Speech unique info	45.66 (29.84)	43.04 (29.36)	46.42 (35.5)	44.07 (34.3)
% Gesture unique info	15.37 (14.92)	14.94 (12.34)	18.18 (18.85)	19.41 (20.07)

Mean values and standard deviations for the various dependent measures by age group and personal common ground condition are listed in Table 2 (for means and SDs by age group and incremental common ground/trial number, see Supplementary materials, Section D). It is interesting to note that for both age groups, the gesture rate (i.e., the gesture frequency normalized by the number of words) was considerably higher in the present task than in Schubotz et al. (2019), where younger adults produced on average 5.89 gestures per 100 words in the CG condition (7.73 in no-CG), and older adults produced on average 5.96 gestures per 100 words in the CG condition (4.88 in no-CG). This difference could be expected, seeing that gestures generally play a more prominent role when talking about spatial topics as opposed abstract or verbal ones (e.g., Feyereisen & Havard, 1999; Lavergne & Kimura, 1987).

3.2. Effects of experimental predictors and cognitive abilities on word count, gesture frequency, and gesture rate

3.2.1. Word count

Word count was predicted by personal common ground, such that fewer words were used in CG than in no-CG trials ($\beta = -8.73$, $SE = 2.06$, $t(154.86) = -4.23$, $p < .001$). Also, there was a significant effect of incremental common ground, such that fewer words were used on later as compared to earlier trials ($\beta = -4.15$, $SE = 0.60$, $t(154.86) = -6.87$, $p < .001$). There was no main effect of age ($\beta = -2.19$, $SE = 6.49$, $t(31.87) = -0.34$, $p = .74$; old vs. young CG: BF = 0.258; old vs. young NCG: BF = 0.224), and no interaction between any of the predictors.

Including verbal WM yielded a significant interaction with incremental common ground, such that participants with higher verbal WM showed a stronger reduction in number of words across trials ($\beta = -1.42$, $SE = 0.60$, $t(150.47) = -2.37$, $p = .02$). Note that participants with lower WM did not fail to reduce but rather started out with a lower number of words on early trials which remained constant across the experiment, while participants with higher WM used a higher number of words on early trials and reduced on later trials (see Fig. 3). None of the other cognitive predictors contributed significantly to the original model.

3.2.2. Gesture count (gesture frequency)

As for word count, the only significant predictors for gesture count were personal common ground ($\beta = -1.40$, $SE = 0.58$, $t(154.77) = 2.42$, $p = .02$) and incremental common ground ($\beta = -0.81$, $SE = 0.17$, $t(154.77) = -4.79$, $p < .001$), with fewer gestures in CG as compared to no-CG trials, and fewer gestures on later trials as compared to earlier trials. There was no main effect for age ($\beta = 0.31$, $SE = 1.60$, $t(31.97) =$

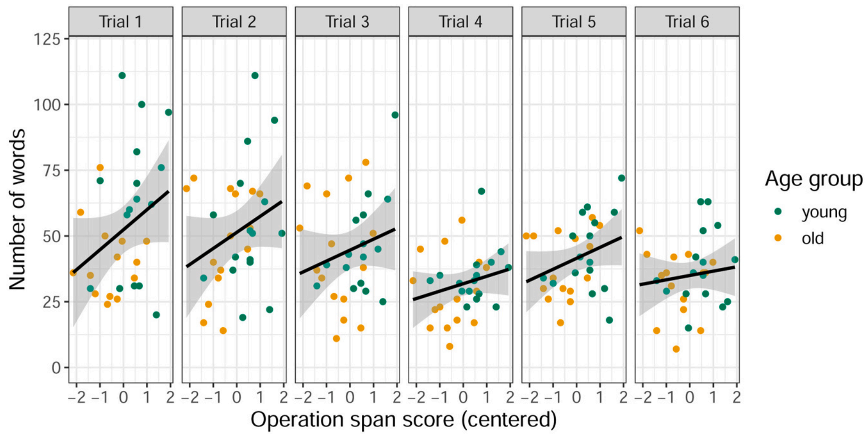

Fig. 3. Interaction effect of verbal WM (z-scored) and incremental common ground on word count, collapsed across CG and no-CG trials. Note that there was no significant effect for age group.

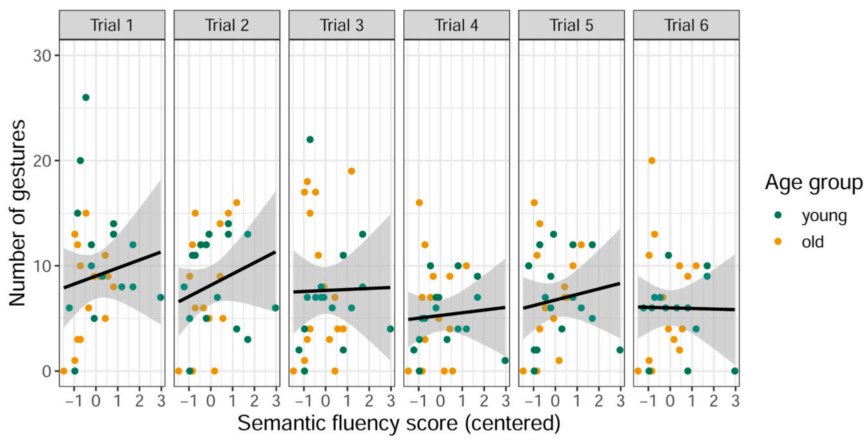

Fig. 4. Interaction effect of semantic fluency (z-scored) and incremental common ground on gesture count, collapsed across CG and no-CG trials. Note that there was no significant effect for age group.

0.20, p = .85; old vs. young CG: BF = 0.215; old vs. young NCG: BF = 0.228), and no interaction with the other predictors.

Including semantic fluency yielded a significant interaction with personal common ground, such that participants with higher semantic fluency showed a stronger reduction in gesture frequency across trials ($\beta = -0.41$, $SE = 0.17$, $t(154.90) = -2.47$, $p = .01$). Note that participants with lower fluency did not fail to reduce but rather started out with a lower number of gestures on early trials which remained constant across the experiment, while participants with higher semantic fluency used a higher number of gestures on early trials and reduced on later trials (see Fig. 4). None of the other cognitive predictors contributed significantly to the original model.

3.2.3. Gesture rate (gestures/100 words)

Gesture rate was not significantly predicted by any of the experimental predictors, age ($\beta = 2.77$, $SE = 3.27$, $t(32) = 0.85$, $p = .40$; old vs. young CG: BF = 0.249; old vs. young NCG: BF = 1.229), personal common ground ($\beta = -0.88$, $SE = 0.97$, $t(160) = -0.91$, $p = .37$), or incremental common ground ($\beta = -0.12$, $SE = 0.28$, $t(160) = -0.42$, $p = .68$). There were also no effects for cognitive predictors.

3.3. Effects of experimental predictors and cognitive abilities on information encoded in speech and in gestures

3.3.1. Number of semantic features encoded in speech and in gestures

The amount of semantic features expressed in speech was predicted only by incremental common ground, such that later trials contained fewer features ($\beta = -0.14$, $SE = 0.05$, $t(160) = -2.70$, $p = .008$). There were no effects for age ($\beta = -0.90$, $SE = 0.56$, $t(32) = -1.17$, $p = .12$; old vs. young CG: BF = 1.740; old vs. young NCG: BF = 1.938) or personal common ground ($\beta = 0.21$, $SE = 0.18$, $t(160) = 1.17$, $p = .25$). Similarly, the amount of semantic features expressed in gesture was predicted only by incremental common ground, such that later trials contained fewer features ($\beta = -0.21$, $SE = 0.06$, $t(154.91) = -3.75$, $p < .001$). There were no effects for age ($\beta = -0.22$, $SE = 0.90$, $t(32) = -0.24$, $p = .81$; old vs. young CG: BF = 0.218; old vs. young NCG: BF = 0.248), or personal common ground ($\beta = 0.33$, $SE = 0.19$, $t(154.91) = 1.75$, $p = .08$). There were no effects for cognitive factors on either measure.

3.3.2. Distribution of information across speech and gestures

The percentage of information expressed uniquely in speech was predicted by incremental common ground, such that more information was encoded uniquely in speech on later as compared to earlier trials ($\beta = 1.95$, $SE = 0.64$, $t(154.90) = 3.04$, $p = .003$). There were no effects for age ($\beta = 0.90$, $SE = 10.13$, $t(32) = 0.09$, $p = .93$; old vs. young CG: BF = 0.216; old vs. young NCG: BF = 0.217) or personal common ground ($\beta = -2.48$, $SE = 2.20$, $t(154.90) = -1.13$, $p = .26$). Similarly, the percentage of information expressed uniquely in gesture was predicted by incremental common ground, such that more information was encoded uniquely in gestures on later as compared to earlier trials ($\beta = 0.96$, $SE = 0.46$, $t(154.79) = 2.12$, $p = .04$). There were no effects for age ($\beta = 3.64$, $SE = 4.69$, $t(31.99) = 0.78$, $p = .44$; old vs. young CG: BF = 0.287;

old vs. young NCG: BF = 0.460) or personal common ground ($\beta = 0.41$, $SE = 1.55$, $t(154.79) = 0.26$, $p = .79$). There were no effects for cognitive factors on either measure.

The percentage of semantic features expressed twice, both in speech and in gestures, was predicted by incremental common ground, such that there was a lower percentage of information encoded twice on later as compared to earlier trials ($\beta = -2.92$, $SE = 0.60$, $t(154.85) = -4.88$, $p < .001$). There were no effects for age ($\beta = -4.54$, $SE = 7.13$, $t(31.99) = -0.64$, $p = .53$; old vs. young CG: BF = 0.271; old vs. young NCG: BF = 0.350) or personal common ground ($\beta = 2.07$, $SE = 2.04$, $t(154.85) = 1.01$, $p = .31$).

Including visuo-sequential WM yielded a significant interaction with personal common ground. In CG trials, participants expressed the same percentage of information twice, in both modalities, regardless of visual WM score. However, in no-CG trials, participants with higher visual WM expressed a higher percentage of information twice, in both modalities, than participants with lower visual WM ($\beta = 4.99$, $SE = 2.01$, $t(155.68) = 2.48$, $p = .01$) (Fig. 5).

3.4. Summary of results

3.4.1. Word and gesture frequency and rate

We found no significant age-related differences in the verbal and gestural behavior of younger and older adults. Both younger and older adults' behavior showed significant effects of personal and incremental common ground: there was a significant reduction in word count and gesture frequency in CG as compared to no-CG trials as well as across the experiment, i.e. going from the first to the final trial. For both age groups, a parallel decrease in both modalities from no-CG to CG trials and across the experiment was indicated by a constant gesture rate.

3.4.2. Information encoded in speech and/or gesture

Again, there were no age-related differences in the amount and distribution of information expressed in speech and in gestures.

The number of features (location, size/shape, orientation) expressed in speech and in gestures decreased across the experiment (incremental common ground). However, there was no effect of personal common ground (CG vs. no-CG trials).

With respect to the distribution of information across the two modalities, the percentage of semantic information encoded uniquely in either of the two modalities increased across the experiment, while the information expressed twice, both in speech and in gesture, decreased across the experiment. That is, we saw a shift across the experiment from encoding information in both modalities, to encoding information only in one single modality.

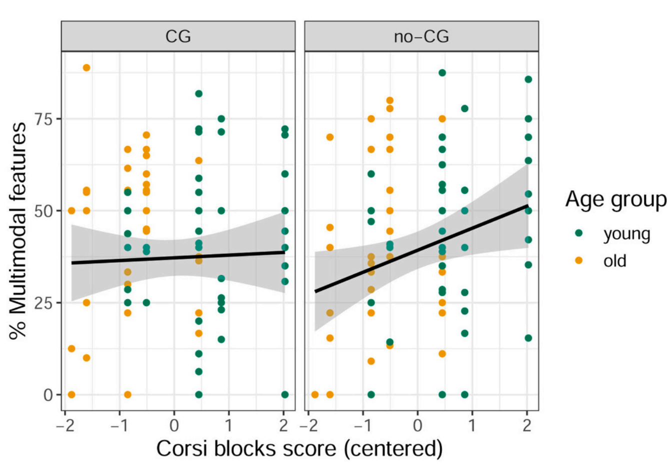

Fig. 5. Interaction effect of visual WM (z-scored) and personal common ground manipulation on percentage of information expressed twice, in both modalities, collapsed across trials. Note that there was no significant effect for age group.

3.4.3. Effects of cognitive predictor variables

Although there were no significant age-related differences on any of these dependent measures and no interaction effects of age group and personal or incremental common ground, we found interaction effects of individual cognitive abilities with the common ground variables. Incremental common ground interacted with verbal WM and with semantic fluency, such that across the experiment, a reduction in word count was more pronounced in individuals with better verbal WM and a reduction in gesture frequency was more pronounced in individuals with higher semantic fluency. Personal common ground interacted with visual WM, such that participants with higher visual WM encoded more information twice, in both modalities, on no-CG trials than participants with lower visual WM.

The Bayes Factors calculated for the comparisons between old and young participants for each of the dependent measures align with the results obtained via the linear mixed effects models including age as a predictor: the large majority of them were less than 0.3, and all of them were lower than 3, thus constituting evidence favoring the H0 or weak evidence for the H1 at best (e.g., Dienes, 2014).

4. Discussion

The present study offers new insights into multimodal recipient design by older and younger adults in a spatial task. Based on previous research, we initially hypothesized that older adults would show less evidence of common ground-based recipient design than younger adults in speech and in gesture in terms of description length and gesture frequency, in terms of information content, and in terms of how the information is distributed across the two modalities. Additionally, we hypothesized that individual differences in cognitive abilities may modulate age-related differences in behavior.

Contrary to our expectations, we found no significant behavioral differences between the two age groups on measures of word and gesture frequency, amount of information expressed in the two modalities or how the information was distributed across the two modalities. Speakers of both age groups adapted their multimodal instructions to our experimentally induced personal and incremental common ground. Rather than by the speakers' age, recipient design in several measures was predicted by individual differences in cognitive abilities. Individual results will be discussed in the following sections.

4.1. Effects of age and personal and incremental common ground on multimodal recipient design

As in Schubotz et al. (2019), we found no age-related differences in overall word count or gesture frequency. This suggests that the relatively lower gesture frequency reported previously for older adults in the visuo-spatial domain (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou et al., 2015) was not attributable to task demands but rather to the lack of a truly interactive setting (see also discussion in Schubotz et al., 2019).

Furthermore, and contrary to what we expected based previous findings (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Schubotz et al., 2019), we also found no age-related differences in verbal and gestural recipient design: both younger and older adults reduced the number of words and of gestures in CG compared to no-CG trials (personal common ground manipulation), as well as across the experiment (incremental common ground manipulation), to the same extent. The parallel decrease in speech and gesture resulted in a constant gesture rate, as has been found in some previous studies (e.g., Campisi & Özyürek, 2013; de Ruiter et al., 2012; Galati & Brennan, 2014; Hilliard & Cook, 2016; see Holler & Bavelas, 2017 for a review). Also, both age groups reduced the amount of information expressed in speech and in gestures across the experiment, i.e., in response to incremental common ground. This reduction in semantic content expressed in speech and in gestures had previously only been observed for manipulations of personal common ground (e.g., Holler & Wilkin, 2009; Schubotz et al., 2019).[1]

Furthermore, we observed a shift in *how* information was distributed across speech and gesture across the experiment. On earlier trials, speakers encoded more information in both modalities, i.e., the semantic features that were expressed in speech were also expressed in gesture and vice versa. Later on, information was more frequently encoded only in speech or, to a lower percentage, only in gesture. Encoding the same piece of information in both modalities is arguably more informative than encoding it only in one modality. Hence, this pattern mirrors the general observation of speaker-gesturers becoming increasingly efficient in terms of their speech and gesture use across the experiment (i.e. as common ground accrues) and provides an additional example of how well the use of the two modalities is coordinated in recipient-designed messages. The absence of any effects of age on this measure, too, suggests that older adults are as skillful as younger adults with respect to the coordination of information across the two modalities in interactive settings. The analyses of speech and gesture "information density", reported in the Supplementary materials, Section A, further corroborate this observation. They are also in line with recent findings by Yoon and Stine-Morrow (2019) who showed that live interaction appears to facilitate older adults' ability to design appropriate verbal utterances for recipients with different knowledge status.

4.2. Interaction effects of cognitive abilities with personal and incremental common ground on multimodal recipient design

Rather than age, we found that individual cognitive variables modulated the extent of verbal and gestural recipient design based on personal and incremental common ground. These findings go beyond previous research on the interplay between cognitive and communicative constraints on speech and gesture use (e.g. Galati & Brennan, 2014; Masson-Carro et al., 2016), as they identify individual differences in specific cognitive variables which influence common ground-based adaptations, rather than inducing an external cognitive load by increasing task demands. It is interesting that these associations surfaced in a visuo-spatial task (see also Long et al., 2018; Wardlow, 2013), but not in a more verbal, narrative task (Schubotz et al., 2019). Possibly, the cognitive measures employed here were better suited to capture the abilities involved in the present task as compared to the narrative task, due to the different cognitive abilities involved.

Verbal WM influenced how strongly speakers reduced the number of words across the experiment, i.e., in response to incremental common ground: individuals with higher verbal WM showed a stronger pattern of reduction than those with lower verbal WM. Presumably, WM resources are needed to update the speaker's discourse model on which information is or is not mutually shared, and to access this information while designing and adapting one's utterances accordingly (see also e.g. Brennan et al., 2010; Horton & Gerrig, 2005; Wardlow, 2013).

Semantic fluency modulated the reduction of gesture frequency in response to incremental common ground: participants with higher semantic fluency showed stronger evidence of gestural adaptations to incremental common ground than those with lower semantic fluency. Potentially, higher semantic fluency, i.e. the efficiency of accessing and retrieving words from existing semantic categories (Martin et al., 1994), allowed speakers to be more flexible in how they used gestures in addition to their verbal message. For example, Hostetter and Alibali (2007) suggest that speakers with high verbal skill may use gestures to make their utterances more communicatively effective, as may also have been the case in the present study.

Finally, we also found that visual WM affected the distribution of information across the two modalities based on personal common ground: individuals with better visual WM encoded more information in both modalities for unknowing addressees than individuals with lower visual WM. This suggests that visual WM, i.e., the ability to store and manipulate visual information, also influences how well speakers can use speech and gesture together for their addressee. We would like to speculate that there might be a mechanism similar to the one proposed for the effects of semantic fluency above: Speakers with higher visual WM may have been more efficient at storing and retrieving the visual information from memory due to their higher spatial skills, and were thus able to use gestures more flexibly in order to tailor their multimodal utterances to their addressees' needs.

Interestingly, our findings are not in line with earlier research suggesting a direct relationship between lower visual or verbal WM (Chu et al., 2014; Gillespie et al., 2014 respectively) and an increase in gesture frequency. It is likely that in other task settings, in which the communicative or interactive function of gestures is less emphasized, the relationship between cognitive abilities and gesticulation manifests itself differently. Yet, note that our findings are based on a relatively small sample. Ideally, future research should replicate these results, using larger sample sizes. Another interesting question for future research to tackle is how explicit and implicit memory of recipients' knowledge affects older adults' *multimodal* audience design in face-to-face interaction, since it has been argued that implicit partner-specific knowledge may help older adults to retain verbal audience design abilities in live interaction.

5. Conclusion

Taken together, our results indicate that like younger adults, older adults were aware of the presence or absence of shared knowledge induced experimentally, i.e., personal and incremental common ground, and could adapt their multimodal utterances accordingly. Additionally, our findings suggest that younger and older adults' common ground-based adaptations were affected by individual differences in cognitive abilities, with higher cognitive performance in verbal and visual WM and semantic fluency allowing for more strongly pronounced recipient design.

Thus, previous findings of age-related deficits in common ground-based recipient design in the verbal (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Schubotz et al., 2019) and gestural domain (Schubotz et al., 2019) do not seem to generalize to the spatial task employed in the present study. First and foremost, by virtue of being spatial, the present task presumably placed different demands on the speech and co-speech gesture production system. The fact that gestures were very prominent during the spatial descriptions may have given speakers the opportunity to "off-load" information onto visual space, thereby freeing up cognitive resources (see Goldin-Meadow et al., 2001; Cook et al., 2012), which then became available for other cognitive operations, like the common ground-based adaptation of utterances. Furthermore, the language used in the present task consisted of a fairly restricted vocabulary, consisting mainly of geometric shape and size attributes and spatial prepositions; this may have additionally decreased the demands of verbal utterance planning, thus leaving more capacity for the cognitive operations involved in recipient design. In addition, the straight-forward nature of the present task presumably reduced age-related differences in task interpretation and communicative goals, which may have contributed to the results obtained by Schubotz et al. (2019). Further, note that the live, interactive nature of the present task alone is unlikely to explain differences with previous studies (cf. Yoon & Stine-Morrow, 2019), since Schubotz et al.'s (2019) task was also based on live interaction.

[1] Interestingly, in the present study, the same amount of verbal and gestural information was expressed in CG as compared to no-CG trials. It appears that regardless of personal common ground condition, speakers always deemed the same amount of information minimally necessary in order to construct the model, which may make this finding specific to our task – after all, the present task was restricted to just four semantic aspects that were relevant for completing the task.

We would like to suggest that this interplay of cognitive *and* social contextual factors determined older adults' communicative behavior, causing the different pattern of results observed in the present task compared to Schubotz et al. (2019). Future research might further explore this possibility, by systematically manipulating the type of cognitive factors involved in a given task, the task difficulty, and the speakers' communicative goals.

In sum, in the present study, we found no evidence that the ability to engage in common ground-based recipient design, both verbally and gesturally, decreases as a function of age. In the spatial instruction task that we employed, both age groups flexibly adapted their speech and co-speech gesture use and the amount of information they expressed in the two modalities according to their addressee's knowledge state in terms of personal and incremental common ground. Importantly, individual differences in verbal and visual WM and semantic fluency modulated the extent of these addressee-based adaptations, such that higher cognitive abilities predicted more strongly pronounced recipient design. We conclude that a combination of situation-specific communicative requirements *and* of cognitive factors determines how younger and older adults speak and gesture in interaction with others.

Declaration of competing interest

The authors do not report any conflict of interest.

Acknowledgements

We would like to thank Nick Woods for his help with video editing, Renske Schilte for her assistance with transcribing the data, and Victoria Poulton and Patrick Hobbelen for their help with the reliability coding.

Funding

LS was supported by an MPG MaxNetAging doctoral stipend during a part of this project. JH was supported by a consolidator grant from the European Research Council (#773079) during a part of this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2022.103690.

References

Adams, C., Smith, M. C., Pasupathi, M., & Vitolo, L. (2002). Social context effects on story recall in older and younger women: Does the listener make a difference? *Journal of Gerontology: Psychological Sciences, 57B*(1), 28–40.
Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation, 5*(4), 307–331.
Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188.
Bates, D., Maechler, M., & Bolker, B. (2017). *lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1–13*. Retrieved from https://cran.r-project.org/web/packages/lme4/.
Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*(4), 394–405.
Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469–489.
Brennan, S., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialogue: Coordinating speaking and understanding. *Psychology of Learning and Motivation, 53*, 301–344.
Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893–900.
Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*, 14–27.
Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive ability and empathy. *Journal of Experimental Psychology: General, 143*(2), 694–709.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39.
Cohen, R. L., & Borsoi, D. (1996). The role of gestures in description-communication: A cross-sectional study of aging. *Journal of Nonverbal Behavior, 20*(1), 45–63.
Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes, 27*(4), 594–610.
de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science, 4*(2), 232–248.
Dienes, Z. (2014). Using bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781.
ELAN. (2016). *(Version 4.9.4) [Computer software]*. Retrieved from. Nijmegen: Max Planck Institute for Psycholinguistics https://tla.mpi.nl/tools/tla-tools/elan/.
Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior, 23*(2), 153–171.
Fussel, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378–391.
Galati, A., & Brennan, S. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.
Galati, A., & Brennan, S. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435–451.
Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition, 132*, 174–180.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516–522.
Healey, H., & Grossman, M. (2016). Social coordination in older adulthood: A dual-process model. *Experimental Aging Research, 42*(1), 112–127.
Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 91–103.
Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground. In R. Breckinridge Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 213–240). Amsterdam: John Benjamins.
Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267–289.
Horton, W. S., & Gerrig, R. (2005). The impact of memory demands on audience design during language production. *Cognition, 96*, 127–142.
Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*(2), 281–290.
Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture, 7*(1), 73–95.
Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8*(3), 339–346.
Kendon, A. (2004). *Gesture: Visible action as utterance*. UK: Cambridge University Press.
Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric bias of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*(2), 46–50.
Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Lecture Notes in Computer Science: 1371. Gesture and Sign Language in Human-computer Interaction. GW 1997* (pp. 23–35). Berlin, Heidelberg: Springer.
Klencklen, G., Després, O., & Dufour, A. (2012). What do we know about aging and spatial cognition?Reviews and perspectives. *Aging Research Review, 11*(1), 123–135.
Kuznetsova, A., Brockhoff, P. B., & Bojesen Christensen, R. H. (2016). *lmerTest: Tests in linear mixed effects models. R package version 2.0-33*. Retrieved from http://cran.r-project.org/web/packages/lmerTest/.
Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia, 25*(4), 689–693.
Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition, 170*, 25–30.
Long, M. R., Rohde, H., & Rubio-Fernandez, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports, 10*, 8214.
Lysander, K., & Horton, W. S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes, 49*(1), 29–60.
Markostamou, I., Coventry, K., Fox, C., & McInnes, L. (2015). Both symbolic and embodied representations contribute to spatial language processing: Evidence from younger and older adults. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society (CogSci 2015)* (pp. 1518–1523). Austin, TX: Cognitive Science Society.
Martin, A., Wiggs, C. L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsychologia, 32*(12), 1487–1494.
Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handel this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition, and Neuroscience, 31*(3), 430–440.
McNeill, D. (1992). *Hand and mind*. Chicago, London: The Chicago University Press.
Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes, 21*, 238–290.

R Core Team. (2015). *R: A language and environment for statistical computing*. Retrieved from. Vienna, Austria: R Foundation for Statistical Computing http://www.R-project.org.

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

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226–231.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.

Saryazdi, R., Bannon, J., & Chambers, C. G. (2019). Age-related differences in referential production: A multiple-measures study. *Psychology and Aging, 34*(6), 791–804.

Schubotz, L., Özyürek, A., & Holler, J. (2019). Age-related differences in multimodal recipient design: Younger, but not older adults, adapt speech and co-speech gestures to common ground. *Language, Cognition and Neuroscience, 34*(2), 254–271.

Theocharopoulou, F., Cocks, N., Pring, T., & Dipper, L. T. (2015). TOT phenomena: Gesture production in younger and older adults. *Psychology and Aging, 30*(2), 245–252.

Underwood, K. (2010). Interactive remembering: Insights into the communicative competence of older adults. *Journal of Aging Studies, 24*, 145–166.

Wardlow, L. (2013). Individual differences in speaker's perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review, 20*(4), 766–772.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics, The Language Archive. http://tla.mpi.nl/tools/tla-tools/elan/.

Yoon, S. O., & Stine-Morrow, E. A. L. (2019). Evidence of preserved audience design with aging in interactive conversation. *Psychology & Aging, 34*, 613–623.