Not everybody has an inner voice: Initial exploration of anendophasia and four behavioral experiments

Johanne S. K. Nedergaard[1] & Gary Lupyan[2]

[1] Department of Linguistics, Cognitive Science and Semiotics, Aarhus University

[2] Department of Psychology, University of Wisconsin-Madison

Author Note

All experiment data, experiment code, and analysis code are available on GitHub: https://github.com/johannenedergaard/anendophasia.

Correspondence concerning this article should be addressed to Johanne S. K. Nedergaard, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark. E-mail: nedergaard@cc.au.dk

Abstract

Inner speech appears to play an important role in a variety of cognitive processes but the existence of people who experience little or no inner speech in their everyday life may limit conclusions we can make about the functions of inner speech. In this exploratory study, we test two groups of participants (one reporting high levels of internal verbal representations and one reporting low levels of internal verbal representations) across four behavioral experiments (verbal working memory, rhyme judgment, same/different judgment, and task switching). These experiments were chosen to test internal verbal representations, verbal self-cuing, and the influence of inner speech on category processing. We found that participants who report low levels of verbal representations performed worse at the verbal working memory and the rhyme judgments and were slower at same/different judgments. Both groups performed equally well on the task switching experiment, however. We propose a new name for the phenomenon of experiencing little to no inner speech - *anendophasia* - and discuss the implications of our findings for the role of inner speech in cognition.

Keywords: inner speech, rhyme judgments, categorization, task switching, verbal working memory, individual differences

Word count: 7927

"To think means to speak to oneself... hence to hear oneself inwardly."

— Immanuel Kant in *Anthropology from a pragmatic viewpoint* (1798)

"He who is capable of thinking with the aid of language alone, has not yet experienced abstract and genuine thinking."

— Eugen Karl Dühring as quoted in Engels (1877)

1 Introduction

Everyone has an inner voice, and most of our waking hours are filled with internal monologue. These are claims that scientists of the mind casually make often (Chella & Pipitone, 2020; Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Loevenbruck, 2014) but the picture is muddied by the recent "discovery" that some people apparently do not experience an inner voice. The topic has received much attention in viral Twitter threads (e.g., @KylePlantEmoji (2020), see Figure 1) as well as in articles such as 'What it's like living without an inner voice' (Soloducha, 2020) and 'People With No Internal Monologue Explain What It's Like In Their Head' (Felton, 2020).

Figure 1. Viral tweet from @KylePlantEmoji about the presence or absence of inner speech. Screenshot from November 17th 2022.

Judging by these accounts, there are important differences in the extent to which people experience an inner voice. Whether these differences in experience result in differences in behavior is still an open question which we hope begin to address in the present article.

If people's reports are accurate and an absence or diminished awareness of inner speech is more widespread than previously thought, this has important consequences for the relationship between language/inner speech and thought. The assumption that everyone has an inner voice has served as a stepping stone for much research into the functions of inner speech – if everyone has it, it must be important. This importance ranges from claims that language constitutes (at least some types of) thought (Bermúdez, 2007; Carruthers, 2002; Clark, 1998; Frankish, 2018; Gauker, 2011; Morin, 2018) to investigations into the connection between inner speech and behavioral control (Alderson-Day & Fernyhough, 2015; Cragg & Nation, 2010; Emerson & Miyake, 2003; Morin, Duhnych, & Racy, 2018). As the two quotes at the beginning of this article suggest, philosophers and psychologists have disagreed about the role of inner speech and language in thinking for centuries. It is possible that these disagreements stem at least in part from the philosophers' and psychologists' subjective experience as this is also the case with aphantasia, the inability to engage in visual imagery (Galton, 1880). With globalization and social media has come the opportunity to discuss these intuitions of a select few thinkers with a much wider range of people whose intuitions we would do well to take seriously.

1.1 Parallels with unsymbolized thinking

Anecdotally, some people with no inner speech report that their thinking takes place largely in the visuospatial modality while another common description is that they 'think in concepts'. What it means to think in concepts without relying on language is contentious but the idea to some extent parallels findings from Descriptive Experience Sampling (DES) (Heavey & Hurlburt, 2008; Russell T. Hurlburt & Akhter, 2006), where participants often report what is called 'unsymbolized thinking' (around 22 % of experience prompts). In such episodes, people feel that they think ‘a particular, definite thought without the awareness of that thought’s being conveyed words, images, or any other symbols’ (Heavey & Hurlburt, 2008, p. 802). Unsymbolized thinking is a slippery phenomenon mostly characterized with negative definitions. For example, Russell T. Hurlburt and Akhter (2008) say that it is experienced as being ‘a thinking, not a feeling, not an intention, not an intimation, not a kinesthetic event, not a bodily event’ (p. 1366). The existence of conscious, unsymbolized thinking poses a challenge to views that hold that language is necessary for conscious thinking, as does the putative existence of people who do not experience inner speech at all. Indeed, proponents of such views tend to deny the existence of unsymbolized thought, claiming that reports may be a result of confabulation (Carruthers, 2009). Either people who report such thoughts are using words or images without being aware of it, or they are not really thinking (e.g. confusing looking at something with considering something). One way of testing the confabulation idea is to test whether the presence or absence of verbal thought has behavioral consequences which is what we did in the series of experiments discussed in the present paper.

1.2 Parallels with aphantasia

That there are differences in subjective reports of inner experience is not a new finding, nor is the idea that such differences may result in subtle behavioral changes. In recent years, a very similar phenomenon to internal verbal experience has gained much attention, namely the presence or absence of visual imagery. In a 2010 article, Zeman and colleagues termed the inability to engage in visual imagery ‘aphantasia’ and reported that two thirds of the participants had difficulties with autobiographical memory as a result of their aphantasia (Zeman et al., 2010). Generally, participants with aphantasia report weak or non-existing ability to visualize ‘in the mind’s eye’ (Dawes, Keogh, Andrillon, & Pearson, 2020; Keogh & Pearson, 2018) and may display poorer visual working memory performance than control participants (Jacobs, Schwarzkopf, & Silvanto, 2018) although this is not always the case

(Keogh, Wicken, & Pearson, 2021). The conflicting findings about consequences of aphantasia in terms of working memory abilities have prompted a discussion of whether aphantasia represents a metacognitive deficit rather than difficulties with mental visual imagery. However, recent findings suggest that a more likely explanation is that people with aphantasia simply use different strategies to solve tasks that would normally require visual imagery. For example, Keogh, Wicken, and Pearson (2021) found that participants with aphantasia performed at the same level as control participants on visual working memory tasks. There were, however, marked differences in the reported strategies used by participants with aphantasia who reported rehearsing patterns verbally or 'using ideas and semantics' to remember the test items. Additionally, performance levels on a number working memory task and a visual working memory task were correlated for participants with aphantasia but not for control participants. This suggests that control participants used different strategies for the two types of tasks (one is traditionally thought to occupy verbal resources while the other is thought to utilize visual working memory resources) while participants with aphantasia may have used similar strategies for the two different tasks. The finding that differences in strategies are likely to mask differences in visualizing ability is important for research in inner speech as well. We might see comparable performance levels due to compensatory strategies that would then mask differences in mental verbalizing abilities.

There are, however, also important differences between aphantasia and lack of inner speech as inner speech cannot be reduced to auditory imagery. First, inner speech involves both auditory and articulatory-motor imagery (Geva, 2018; Perrone-Bertolotti et al., 2014). Second, while internal language addressed to oneself is often experienced as having phonological features - one of the reasons people often perceive it as speech (Langland-Hassan, 2018) - it does not necessarily. For example, inner speaking and inner signing have something in common as ways of producing internal language addressed to oneself that has nothing to do with auditory imagery (McGuire et al., 1997).

1.3 How widespread might a lack of inner speech be?

It is difficult to assess the prevalence of inner speech in the general population but there are nevertheless several sources for making educated guesses. People with aphantasia often report difficulties with imagery in other modalities as well (Dawes et al., 2020) and so we might expect a similar incidence rate for lack of inner speech. If we take this as a starting point, we could see little to no inner speech in around 2-4 % of the population (Dance, Ipser, & Simner, 2022; Faw, 2009). Alternatively, we could look at what proportion of children engage in private speech (93 % according to Winsler, De Leon, Wallace, Carlton, and Willson-Quayle (2003)) or how many adult participants report never engaging in self-talk in large-scale questionnaire studies (e.g. less than 10 % in Nedergaard, Christensen, and Wallentin (2021)). Regardless, people with no inner speech appear to be in a small minority, and the phenomenon is likely to be continuous rather than dichotomous.

1.4 Which behavioral consequences of less inner speech would we expect?

Say that we take people's claims seriously when they say that they experience no inner speech. Given the myriad of theories on the role that language plays in cognition, we would expect them to also think differently. For example, we might believe that inner speech is recruited for (or actually constitutes) self-reflection and metacognition (Morin, 2018; Clark, 1998), and so people with less inner speech should perform less well on metacognitive tasks. If we find both that they are right about not having inner speech and that they perform equally well, then we might conclude that inner speech is probably not crucial for metacognition. This line of reasoning hinges on finding a good way to test whether people are right when they claim that they experience no inner speech and finding tests of cognitive abilities that either could only be solved with inner speech or would allow us to differentiate between verbal and non-verbal solutions/strategies (despite them possibly resulting in comparable performance).

1.5 The present study

To test whether participants are accurate about their internal verbal representations, we use a rhyme judgment task (Geva, Bennett, Warburton, & Patterson, 2011; Langland-Hassan, Faries, Richardson, & Dietz, 2015) where participants see two images and have to judge whether the associated words rhyme or not. Presumably, this would require them to internally verbalize. Importantly, we need to include both orthographic rhymes (such as 'boat' and 'moat') and non-orthographic rhymes (such as 'sleigh' and 'hay') as participants could otherwise make rhyme judgments by visualizing the orthographic representations of the words. In functional neuroimaging studies, rhyme judgment tasks are associated with activation in language production areas such as the left inferior frontal gyrus and the inferior parietal lobe (Hoeft et al., 2007; Lurito, Kareken, Lowe, Chen, & Mathews, 2000; Owen, Borowsky, & Sarty, 2004; Paulesu, Frith, & Frackowiak, 1993; Poldrack et al., 2001; Pugh et al., 1996). Taking inspiration from aphantasia research on visual working memory, it also seems reasonable to focus on verbal working memory in the case of people with little to no inner speech. In particular, we might expect difficulties with verbal working memory tasks requiring a high degree of phonological precision (Jacobs et al., 2018). For this, we use verbal working memory tests with orthographically similar sets, phonologically similar sets, and control sets.

There is robust evidence that inner speech is generally recruited for behavioral control in task switching paradigms where participants have to switch between different task rules (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Miyake, Emerson, Padilla, & Ahn, 2004). Here, participants for example have to switch between adding and subtracting numbers while also performing simultaneous tasks designed to interfere (see Nedergaard, Wallentin, and Lupyan (2022) for a systematic review of the verbal interference literature). People who do not habitually use inner speech might resort to other means of self-cueing in such cases. Given the robust literature on inner speech being involved in task switching, we also test participants' ability to switch between simple addition and subtraction problems. This helps shed light on the consequences of (lack of) inner speech for behavioral control.

Aside from behavioral control, there is also empirical evidence that language influences thought on a more structural level through imposing categories on for example the color spectrum (Gilbert, Regier, Kay, & Ivry, 2006, 2008; Winawer et al., 2007). Given that even people who report no inner speech grow up using language and learning categories through language, it is perhaps more of an open question whether we would expect them to be influenced by categories to the same extent as people with a lot of inner speech. Related to this idea that language structures the way we think about categories, we also test participants' ability to detect simple visual differences as well as categorical differences between line drawings of cats and dogs.

2 Methods

2.1 Participants

We contacted participants who had previously completed the Internal Representations Questionnaire (Roebuck & Lupyan, 2020) and recruited ones with the lowest and highest scores on verbal representation. This study received ethical approval from XXXX. ((**Gary?**) how exactly was this done?) We will use the shorthand 'high-verbal' and 'low-verbal' to refer to these two groups although it should be noted that these just refer to their inclination or propensity to engage in verbal mental imagery and not to their general linguistic abilities. All four experiments were conducted using custom-written software with the JavaScript package jsPsych version 6 (De Leeuw, 2015).

2.2 Method: Verbal working memory

2.2.1 Materials. We used the same three word sets as Baddeley (1966): One set contained words that were phonologically similar but not orthographically similar ('bought', 'sort', 'taut', 'caught', and 'wart'), one set contained words that were orthographically similar but not phonologically similar ('rough', 'cough', 'through', 'dough', 'bough'), and one set was a control set ('plea', 'friend', 'sleigh', 'row', 'board').

2.2.2 Procedure. Participants received the following instructions: 'During each trial, you will see five words presented in sequence 1 second each. You will have to remember the words and the order they were presented in as you will be asked to reproduce them in the right order afterwards.' There was a blank screen for 2000 ms between each word presentation. First, participants performed two practice trials with full feedback (correct/incorrect and the stimulus words – drawn from a different set than the ones used in the real experiment – shown in order). Then, participants performed 24 trials with eight trials from each of the three word sets. The order of both set type and words within a trial were randomized. There was no limit to how long participants could spend on reproducing the words on a given trial. See Figure 2.

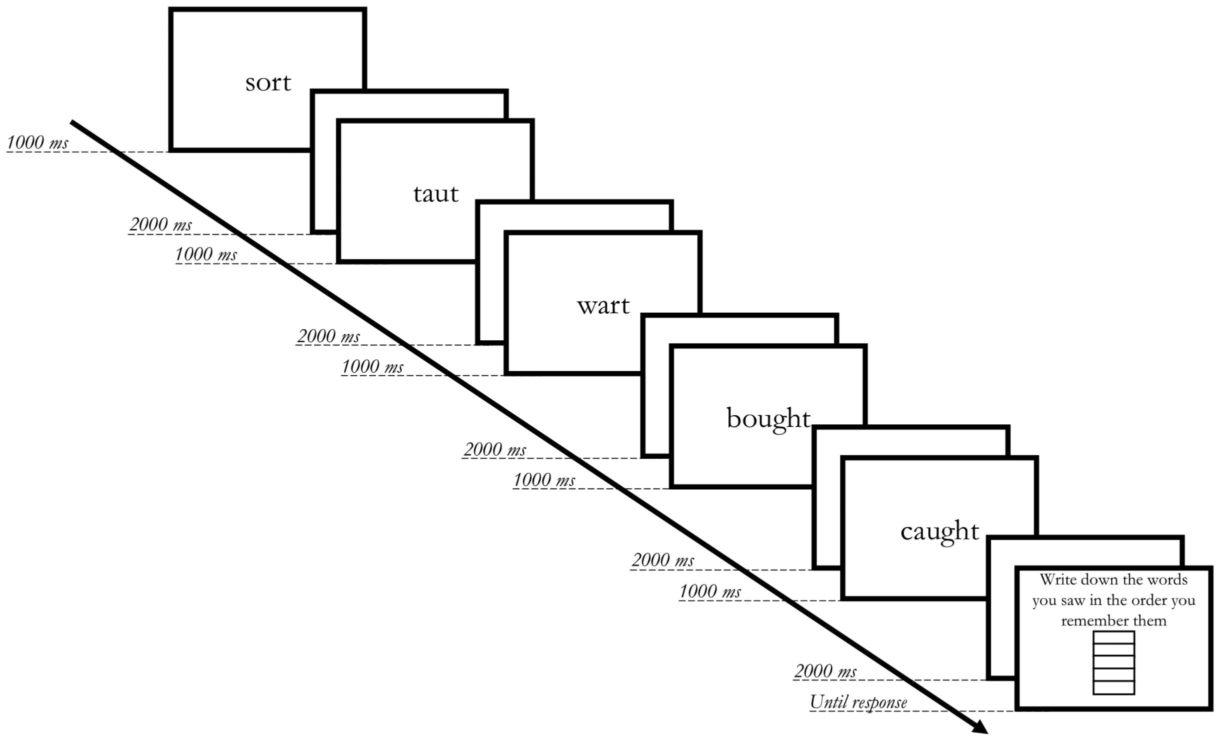

Figure 2. A sketch of the procedure in the verbal working memory experiment. In this example, the words are drawn from the phonological similarity set

2.3 Method: Rhyme judgments

2.3.1 Materials. We constructed a set of rhyme pairs with 20 orthographical pairs (e.g. 'sock' and 'clock') and 20 non-orthographical pairs (e.g. 'drawer' and 'door'). See Supplementary Materials for the full set of images, associated words, and name agreement scores. The images were selected from the MultiPic database (Duñabeitia et al., 2018) and from Snodgrass and Vanderwart (1980).

2.3.2 Procedure. Participants received the following instructions: 'You will see two images at a time and have to judge whether the names of the items rhyme or not. For example, if you see a picture of a LAMP and a picture of a CAMP, you should respond that they rhyme (press UP arrow). If you see a picture of a BEAR and a picture of a CUP, you should respond that they do not rhyme (press DOWN arrow). All the words are short (one syllable). Please make the judgments as quickly and accurately as possible.' Participants first performed four practice trials with correct/incorrect feedback – they did not receive feedback for the remaining trials. Between each rhyme judgment trial, the screen showed a central fixation cross for either 250, 500, 750, or 1000 ms. It then showed two square black frames for 500 ms to control spatial attention – the two images then appeared simultaneously in the two squares. Participants had 5000 ms to respond to each trial and performed a total of 60 rhyme judgments in randomized order (20 orthographical rhymes, 20 non-orthographic rhymes, and 20 no-rhyme control trials). See Figure 3.

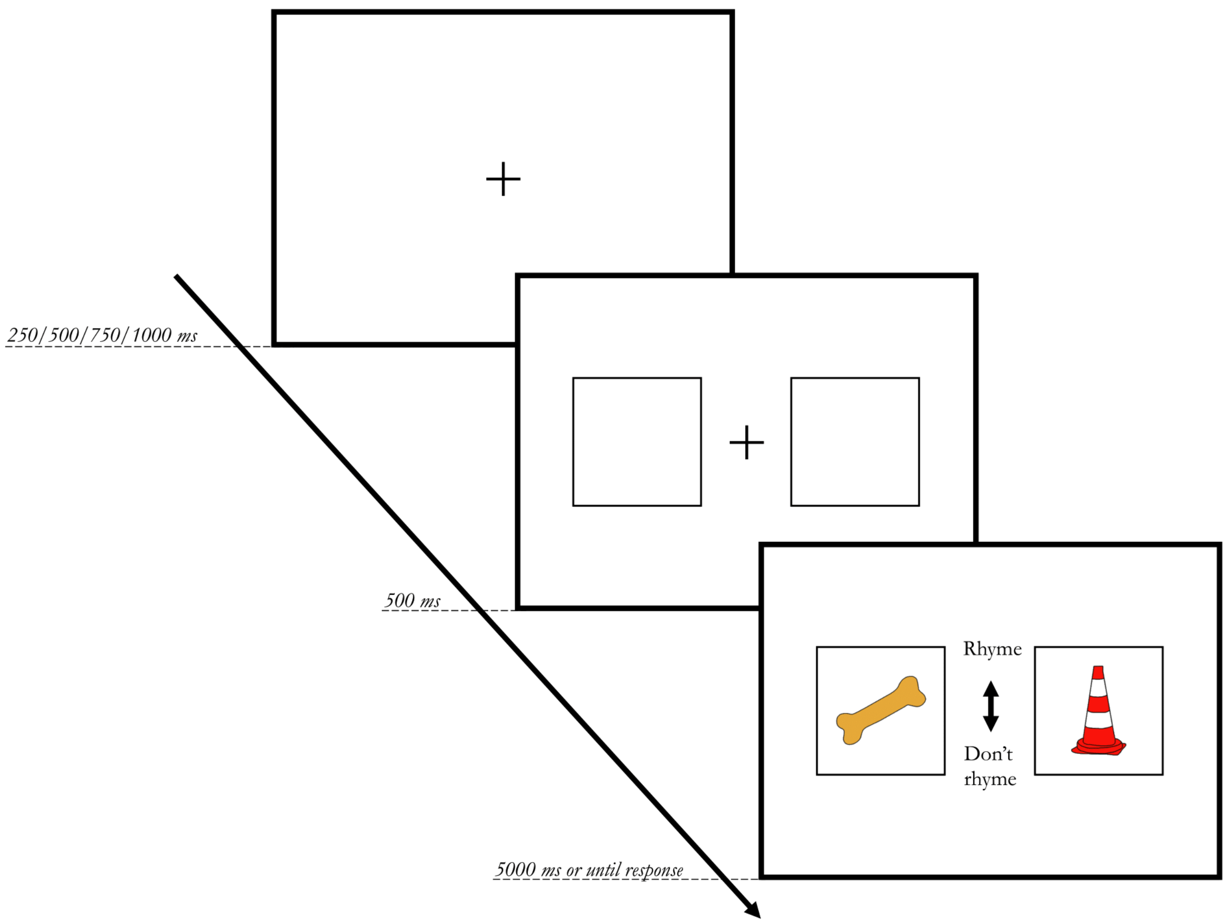

Figure 3. A sketch of a rhyme judgment trial. The stimuli here exemplify an orthographic rhyme – 'bone' and 'cone' – and the correct answer would therefore be 'Rhyme'.

2.4 Method: Task switching

2.4.1 Materials.

For each of the five experimental conditions, we used 30 randomly selected integers between 13 and 96 as prompts to make sure that all correct results were two-digit positive numbers.

2.4.2 Procedure.

There were five conditions in this experiment: (1) blocked addition (2) blocked subtraction, (3) alternating between addition and subtraction with operation marked by color cue (red/blue); (4) alternation marked with a symbol cue (+/-); (5) alternation without external cue requiring participants to remember which operation they just did.. Participants started with either the blocked subtraction or the addition condition (counterbalanced) and then proceeded to the switching conditions (counterbalanced). For each condition, participants first solved 10 problems with correct/incorrect feedback (including feedback specific to whether the arithmetic or the operation or both were incorrect) and then 30 problems without feedback. On a given trial, participants saw a prompt number and had to either add or subtract 3 and type their answer into a text box. In the switching conditions, a response counted as correct if it was the correct arithmetic and if the operation was switched from the previous trial (from addition to subtraction or vice versa). See Figure 4.

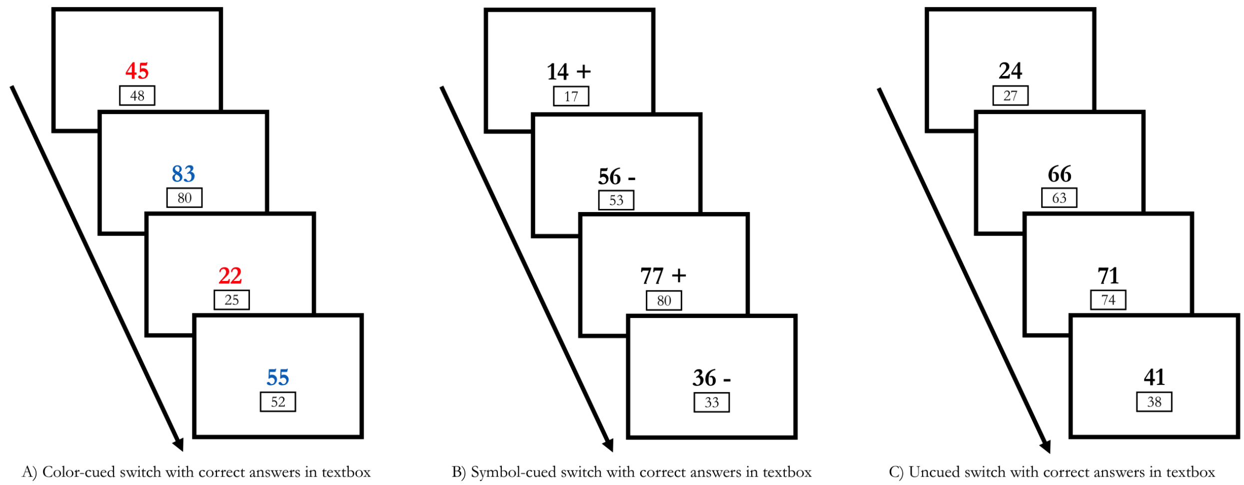

Figure 4. A sketch of the three switched conditions in the task switching experiment. Figure A shows four color-cued switch trials with correct answers, Figure B shows four symbol-cued switch trials with correct answers, and Figure C shows four uncued switch trials with correct answers.

2.5 Method: Same/different judgments

2.5.1 Materials. This experiment used three different black silhouettes of cats and three different black silhouettes of dogs (see Figure 5).

Figure 5. The black silhouettes of cats and dogs used in the same/different judgment experiment.

2.5.2 Procedure. There were two conditions in the experiment: a category judgment condition and an identity judgment condition. In the category judgment condition, participants were instructed to press the UP arrow key if the two animals belonged to the same category (either cat or dog) and the DOWN arrow key if they did not. In the identity judgment condition, participants were instructed to press the UP arrow key if the two animals were completely identical (e.g., same silhouette of same dog) and the DOWN arrow key if they were not. See Figure 6. On each trial, participants first saw a fixation cross for 750 ms, then four empty square frames around the fixation cross for 500 ms to prompt participants' spatial attention. The silhouette images appeared one at a time with a 300 ms delay between them in two out of four random positions around a fixation cross in the center of the screen. After the keyboard response, the screen was blank for 300 ms. Participants received feedback throughout but only for incorrect trials. They completed 100 trials in the category judgment condition and 100 trials in the identity judgment condition (half 'same' and half 'different').

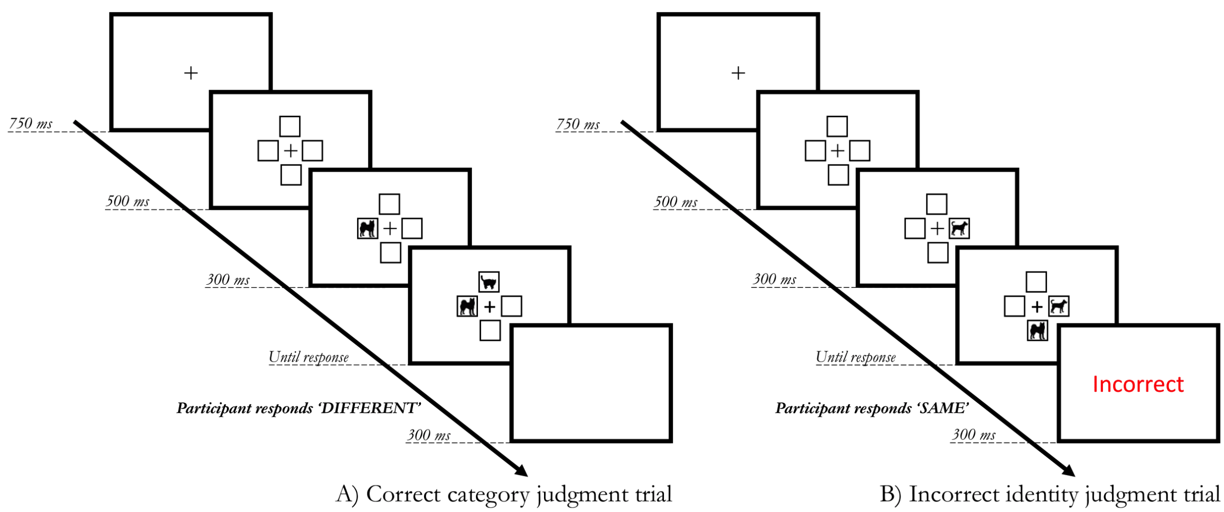

Figure 6. A sketch of the two conditions of the category judgment experiment. On Figure A, we see a correct category judgment trial where the participant responds that the cat and dog silhouettes represent different animals. On Figure B, we see an incorrect identity judgment trial where the participant responds that the two dogs are identical.

2.6 Method: Questionnaire

After completing the four experiments, participants answered the following custom questions. They also completed the Varieties of Inner Speech Questionnaire (VISQ)

(Alderson-Day, Mitrenga, Wilkinson, McCarthy-Jones, & Fernyhough, 2018).

Question	Options
If you have to ask a question in front of an audience, which of these best describes what you typically do?	I rehearse in my mind the exact phrasing of what I am going to ask (5) I rehearse in my mind some of what I am going to ask before asking it (4) I think of a question I want to ask and just ask it (3) Other (2) I'm never in a position to ask questions in front of an audience (1)
How often do you experience trouble focusing on a face-to-face conversation you are having because of a conflicting conversation happening in your mind at the same time?	Never (1) Rarely (2) Sometimes (3) Often (4) Always (5)
How often do you have songs stuck in your head?	Multiple times a day (5) A few times a week (4) A few times a month (3) A few times a year (2) Never (1)
If you had to recall a short conversation about a specific topic that you had yesterday with a friend, how easily can you recall the exact words your friend said?	I can easily recall it. If I wrote it down and matched to a recording of the conversation, there'd be an almost perfect match (5) I remember the topic and remember much of what was said. If I matched it to a recording of the conversation, a lot would match up. (4) I remember the topic, but remember only a few of the specific words/sentences. (3) I remember the topic, but can't remember any of the specifics. (2) Other (1)

(continued)

Question	Options
If you had to recall a short conversation about a specific topic that you had yesterday with a friend, how easily can you recall the exact words you said?	I can easily recall it. If I wrote it down and matched to a recording of the conversation, there'd be an almost perfect match (5) I remember the topic and remember much of what was said. If I matched it to a recording of the conversation, a lot would match up. (4) I remember the topic, but remember only a few of the specific words/sentences. (3) I remember the topic, but can't remember any of the specifics. (2) Other (1)
When you recall a conversation like the one you were thinking about for the last 2 questions, do you hear the words in your mind?	It's just like I'm hearing the conversation again. (4) I hear a condensed version (e.g. only some words). (3) I hear something but I can't describe it. (2) I can't hear it, but I can still recall it. Please briefly say something about how you are recalling it. (1)
Can you "sing along" to music without singing out loud?	Yes - definitely (4) Yes - somewhat (3) No - but I can imagine how others can do it (2) No - I can't imagine how anyone could do this (1)
If you can "sing along" to music without singing out loud, to what extent does this feel like regular thinking?	Not at all (1) Mostly different from regular thinking (2) Neutral (3) Mostly similar to regular thinking (4) Exactly like regular thinking (5) I can't sing along without singing out loud (6)
If you imagine someone else speaking, how do you experience their voice?	I hear what they say in their voice. (4) I hear what they say but in my own voice. (3) I hear the words but I can't tell whose voice it is. (2) I don't "hear" anything, I imagine it by... (please specify) (1)
Many people feel that a lot of their thinking, planning, and decision-making takes place in the form of a conversation with themselves. They describe that when they think, they hear words in their mind. Other people don't have this experience and instead say that they "think in ideas". Is your experience more like the first or the second?	More like a conversation (2) More like "thinking in ideas". Can you elaborate or give an example of what this means to you? (1)

(continued)

Question	Options
To what extent do you agree with this statement: 'It is generally difficult and takes effort to express in words how I think and feel'.	Strongly agree (1) Agree (2) Neither agree nor disagree (3) Disagree (4) Strongly disagree (5)
Do you think it is stressful and annoying to have an inner monologue?	Yes, very (3) Maybe a little (2) No, I don't think so (1)
In books and movies, we often see characters talking to themselves at length. How much do you think this reflects real life?	It's just for the viewer/reader's benefit (1) It might be like real life but mostly for the viewer's/reader's benefit (2) It's exactly like real life (3)
Have you been diagnosed with dyslexia or another reading disorder?	Yes, officially diagnosed (1) Yes, self-diagnosed (2) No, never (3)
Do you ever revise past conversations in your mind (i.e. think of a better comeback, a way of phrasing what you wanted to say)?	Never (1) Rarely (2) Sometimes (3) Often (4) Very often (5)
Do you ever rehearse a conversation before you have it in real life where you simulate what you will say and how the other person will respond?	Never (1) Rarely (2) Sometimes (3) Often (4) Very often (5)
Imagine you are lying in bed with your eyes closed trying to fall asleep. Is your inner experience then...	Primarily verbal (you "hear" or "speak" words and sentences in your mind) (1) Primarily visual (you "see" situations, objects, people etc. in your mind) (2) Primarily about sensory awareness (what you are hearing, smelling, and feeling in the moment) (3) Primarily emotional (4) An even mix of verbal, visual, sensory, and emotional (5) My inner experience in that situation does not have a specific "format" (6)

(continued)

Question	Options
To what extent do you agree with this statement: "I do not know why I do some of the things that I do."	Strongly disagree (1) Disagree (2) Neither agree nor disagree (3) Agree (4) Strongly agree (5)
To what extent do you agree with this statement: “I am a firm believer in thinking things through.”	Strongly disagree (1) Disagree (2) Neither agree nor disagree (3) Agree (4) Strongly agree (5)
To what extent do you agree with this statement: “I like to act on a whim.”	Strongly disagree (1) Disagree (2) Neither agree nor disagree (3) Agree (4) Strongly agree (5)
For each scale, please indicate what percent of people you know you think have each of these three experiences: - Experience their thoughts in the form of a conversation with themselves - Can see vivid images in their mind’s eye - Hear words in their mind’s ear when they silently read	No one (0%) to Everyone (100%)

2.7 Data analysis

We used R (Version 4.1.3; R Core Team, 2022) and the R-packages *corrplot2021* (Wei & Simko, 2021), *cowplot* (Version 1.1.1; Wilke, 2020), *data.table* (Version 1.14.0; Dowle & Srinivasan, 2021), *dplyr* (Version 1.0.7; Wickham, François, Henry, & Müller, 2021), *forcats* (Version 0.5.1; Wickham, 2021a), *Formula* (Version 1.2.4; Zeileis & Croissant, 2010), *ggforce* (Version 0.3.3; Pedersen, 2021), *ggplot2* (Version 3.3.5; Wickham, 2016), *ggpubr* (Version 0.4.0; Kassambara, 2020), *Hmisc* (Version 4.5.0; Harrell Jr, Charles Dupont, & others., 2021), *kableExtra* (Version 1.3.4; Zhu, 2021), *lattice* (Version 0.20.45; Sarkar, 2008), *lme4* (Version 1.1.27.1; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *Matrix* (Version 1.4.0; Bates & Maechler, 2021), *optimx* (Nash, 2014; Version 2021.10.12; Nash & Varadhan, 2011), *papaja* (Version 0.1.1; Aust & Barth, 2022), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version 2.1.1; Wickham, Hester, & Bryan, 2021), *stringr* (Version 1.4.0; Wickham, 2019), *survival* (Version 3.2.13; Terry M. Therneau & Patricia M. Grambsch, 2000), *tibble* (Version 3.1.6; Müller & Wickham, 2021), *tidyr* (Version 1.1.4; Wickham, 2021b), *tidyverse* (Version 1.3.1; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *trackdown* (Version 1.1.1; Kothe, Callegher, Gambarota, Linkersdörfer, & Ling, 2021), and *tufte* (Version 0.12; Xie & Allaire, 2022) for all our analyses. In all linear mixed models, we modelled participant as random intercepts as well as random intercepts for all within-subjects factors (e.g. condition with repeated measures). For the statistical models predicting reaction time, we log-transformed the dependent variable to prevent issues with non-normal distributions. For the statistical models predicting accuracy, we constructed binomial generalized linear mixed models. All the plots visualize categorical differences between the two groups while all the statistical models use verbal score (average score on the verbal representation items on the Internal Representations Questionnaire) as a continuous predictor.

2.7.1 Exclusion criteria. We excluded 10 participants for responding randomly, missing at least one out of the four experiments, or otherwise not complying with task instructions. Our final sample included 47 participants with high verbal representation scores on the IRQ and 46 participants with low verbal representation scores.

3 Results

3.1 Verbal working memory

3.1.1 Descriptive statistics by group: Verbal working memory.

High-verbal participants recalled more words correctly. This advantage was evident both when we scored only correctly ordered responses as correct as well as when we scored correctly recalled items regardless of their position (see Table 2 and Figure 7).

Table 2

Descriptive statistics by group in the verbal working memory experiment.

Group	Word set	Score (item and position)	95% CI score (item and position)	Score (position indifferent	95% CI score (position indifferent)
High-verbal	Control set	4.19	0.13	4.51	0.08
High-verbal	Orthographical similarity set	3.72	0.14	4.18	0.10
High-verbal	Phonological similarity set	3.43	0.16	4.11	0.10
Low-verbal	Control set	3.69	0.15	4.17	0.11
Low-verbal	Orthographical similarity set	3.52	0.15	4.10	0.11
Low-verbal	Phonological similarity set	3.02	0.15	3.81	0.11

3.1.2 Statistical models: Verbal working memory. We conducted two linear mixed models of original word set (phonologically similar, orthographically similar, and control set) and verbal score predicting either memory performance with both correct word and correct position or memory performance with correct word regardless of position. Both models included random intercepts by participant for the three word sets. For memory performance requiring both accurate word and position, both the set with phonologically similar words and the set with orthographically similar words were more difficult than the control set (phonological set: β = -0.72; SE = 0.08; t = -8.84; $p < .001$; orthographical set: β = -0.33; SE = 0.08; t = -3.98; $p < .001$). A higher verbal score was associated with increased memory performance (β = 0.27; SE = 0.10; t = 2.60; p = 0.01). There were no interaction effects (all $p > 0.19$). The same pattern was found when the correct word in any position counted as correct: The phonologically similar set was more difficult than the control set (β = -0.37; SE = 0.07; t = -5.51; $p < .001$) and so was the orthographically similar set (β = -0.21; SE = 0.06; t = -3.35; p = 0.00). A higher verbal score was associated with increased memory performance (β = 0.19; SE = 0.08; t = 2.57; p = 0.01). There were

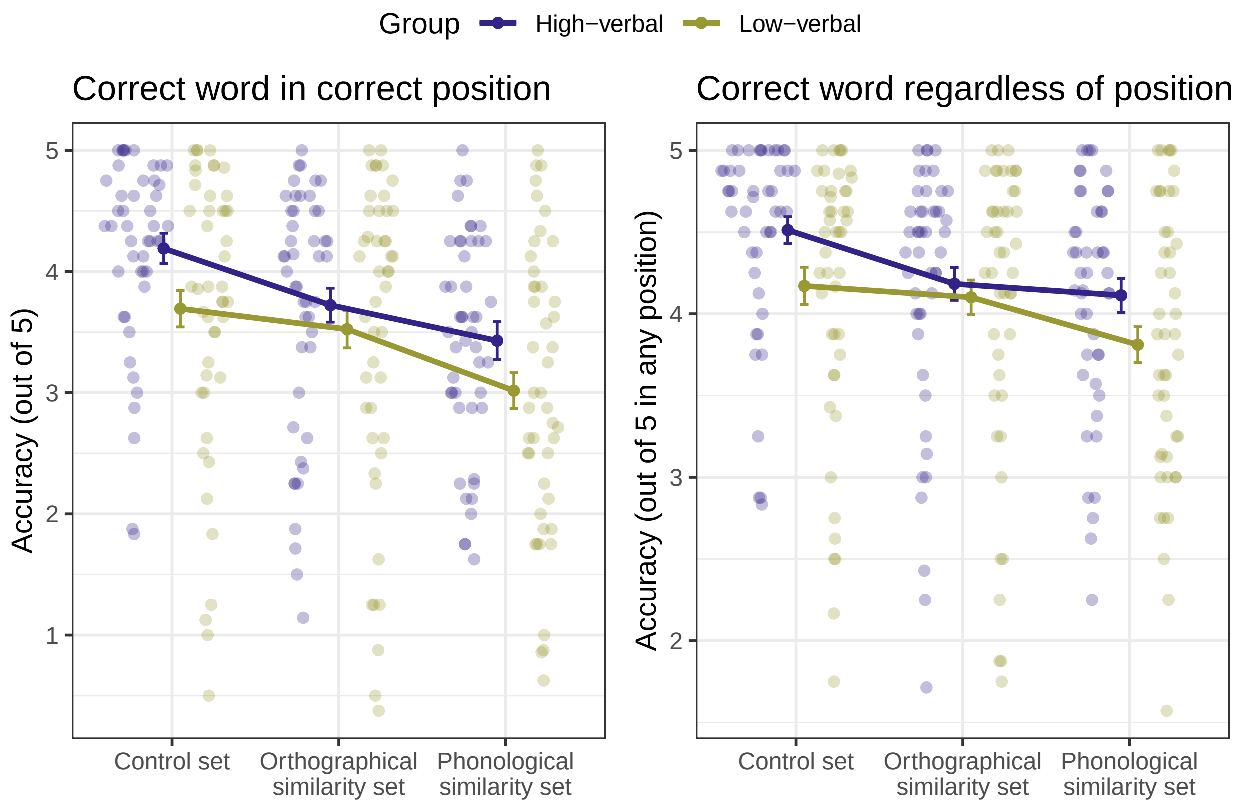

Figure 7. Score on the verbal working memory task by word set. Error bars indicate 95 % confidence intervals.

no interaction effects (all p > 0.10).

3.1.3 Strategies: Verbal working memory. We were interested in whether participants said the words out loud to help them remember them. We asked about this at the end of the experiment. A chi-squared test showed that there was no significant difference between how many high-verbal participants (10 out of 47) and how many low-verbal participants (13 out of 46) reported that they had said the words out loud ($\chi^2(1) = 0.29$, $p = 0.59$). Nevertheless, the effect of doing so was interestingly different for the two groups as can be seen in Figure 8.

The difference between the two groups' memory performance disappears when they report that they said the words out loud to help them remember. Doing so helps low-verbal participants but makes no difference for high-verbal participants. Participants gave some

Figure 8. Verbal working memory performance by whether participants reported talking out loud to help them remember or not. Error bars indicate 95 % confidence intervals.

interesting alternative strategies in response to the free answer question about strategies:

High-verbal group

- Remembering the order of the first letters once the words were familiar (e.g. c, b, t, r, d for ‘cough’, ‘bough’, ‘through’, ‘rough’, ‘dough’). One participant reported this.
- Finding a cadence/melody and using this to repeat the words.
- Chunking.
- Hand and body gestures.
- Creating a story or a sentence with the words in order (both visual and verbal). This one was the most common strategy.

Low-verbal group

- Remembering the order of the first letters once the words were familiar (e.g. c, b, t, r, d for 'cough', 'bough', 'through', 'rough', 'dough'). This strategy was much more common for the low-verbal group than for the high-verbal group.
- Form a story or a narrative. This was a less common strategy than remembering the first letters.

3.2 Rhyme judgments

We excluded five rhyming pairs as they had below-chance performance on average for at least one group. These pairs were bin/chin, cab/crab, rake/cake, wave/cave, and park/shark.

3.2.1 Descriptive statistics by group: Rhyme judgments. Here is a table of accuracy and reaction time for the two groups (high- and low-verbal) across types of rhyming trials.

As can be seen in Table 3, high-verbal participants were generally both faster and more accurate than low-verbal participants on all three types of trials. See also Figure 9.

3.2.2 Statistical models: Rhyme judgments. Both the model predicting reaction time and the model predicting accuracy included random intercepts by participant for the within-subjects variable rhyme type. All predictors were centered. A model of verbal score, rhyme type, and name agreement for the first image predicting log-transformed reaction time showed no main effect of verbal score (β = -0.01; SE = 0.02; t = -0.63; p = 0.53), but it did find a significant effect of no-rhyme trials being slower than other trials (β = -0.04; SE = 0.01; t = -5.90; p < .001) and a significant effect of higher name agremeent being associated with faster reaction times (β = -0.04; SE = 0.00; t = -8.44; p < .001). There were no significant interactions between rhyme type and verbal score (p = 0.56). Another model of verbal score, rhyme type, and name agreement for the first image predicting accuracy showed that no-rhyme trials were easier than other trials (β = -0.95; SE = 0.09; z = -11.03; p

Table 3

Descriptive statistics on rhyming accuracy and reaction time by group and by rhyme type.

Group	Type of rhyme trial	Reaction time (ms)	95% CI (reaction time)	Accuracy	95% CI (accuracy)
High-verbal	Non-orthographical rhyme	1852.66	51.47	82.77	2.86
High-verbal	No rhyme	1930.79	53.26	97.52	1.36
High-verbal	Orthographical rhyme	1719.41	54.99	91.21	2.48
Low-verbal	Non-orthographical rhyme	1970.28	53.85	76.20	3.21
Low-verbal	No rhyme	2024.48	60.47	93.84	1.87
Low-verbal	Orthographical rhyme	1858.94	60.38	83.62	3.22

$< .001$) and that a higher verbal score was associated with a higher likelihood of responding accurately ($\beta = 0.30$; SE = 0.10; z = 3.02; p = 0.00). It also showed that trials with images with higher name agreement were significantly easier ($\beta = 0.13$; SE = 0.04; z = 3.29; p = 0.00). There was no significant interaction between rhyme type and verbal score (p = 0.40).

3.2.3 Strategies: Rhyme judgments. We were interested in whether participants said the words out loud to make the rhyme judgments and so we included this as a question at the end of the rhyming experiment. A chi-squared test showed that there was no significant difference between how many high-verbal participants (23 out of 47) and how many low-verbal participants (21 out of 46) reported that they had said the words out

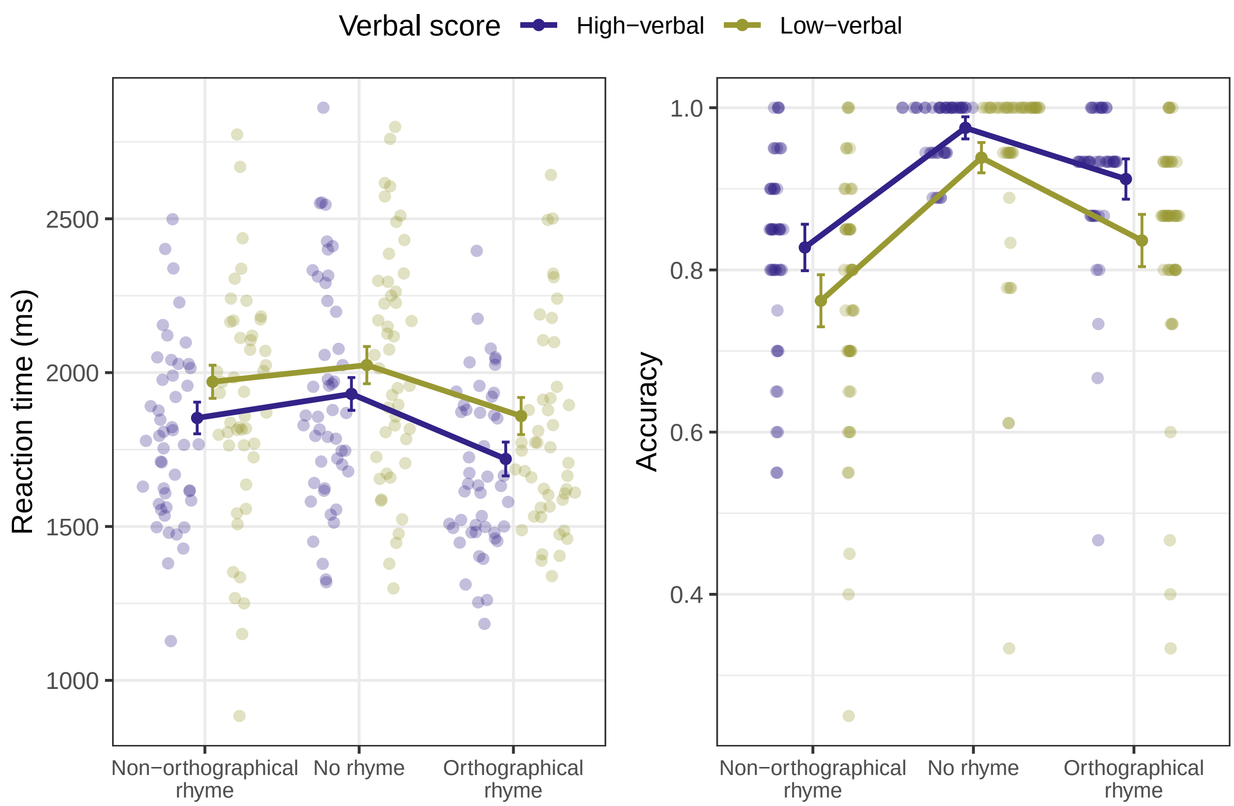

Figure 9. Reaction time and accuracy across groups by rhyme type. Error bars indicate 95 % confindence intervals.

loud ($\chi^2(1) = 0.01$, p = 0.91). Nevertheless, the effect of doing so was interestingly different for the two groups as can be seen in Figure 10.

For both reaction time and accuracy, saying the words out loud diminished the difference between the two groups. This suggests that this was the strategy that high-verbal participants used in their heads - indeed, this was the most common strategy provided by the participants (from both groups) who chose to answer the free answer about strategy. There were no other notable strategies from the free answers.

3.3 Task switching

We excluded trials over 10 seconds (0.5 % of trials). We also recalculated the accuracy measure so that any trial in the three switch conditions where participants in fact switched

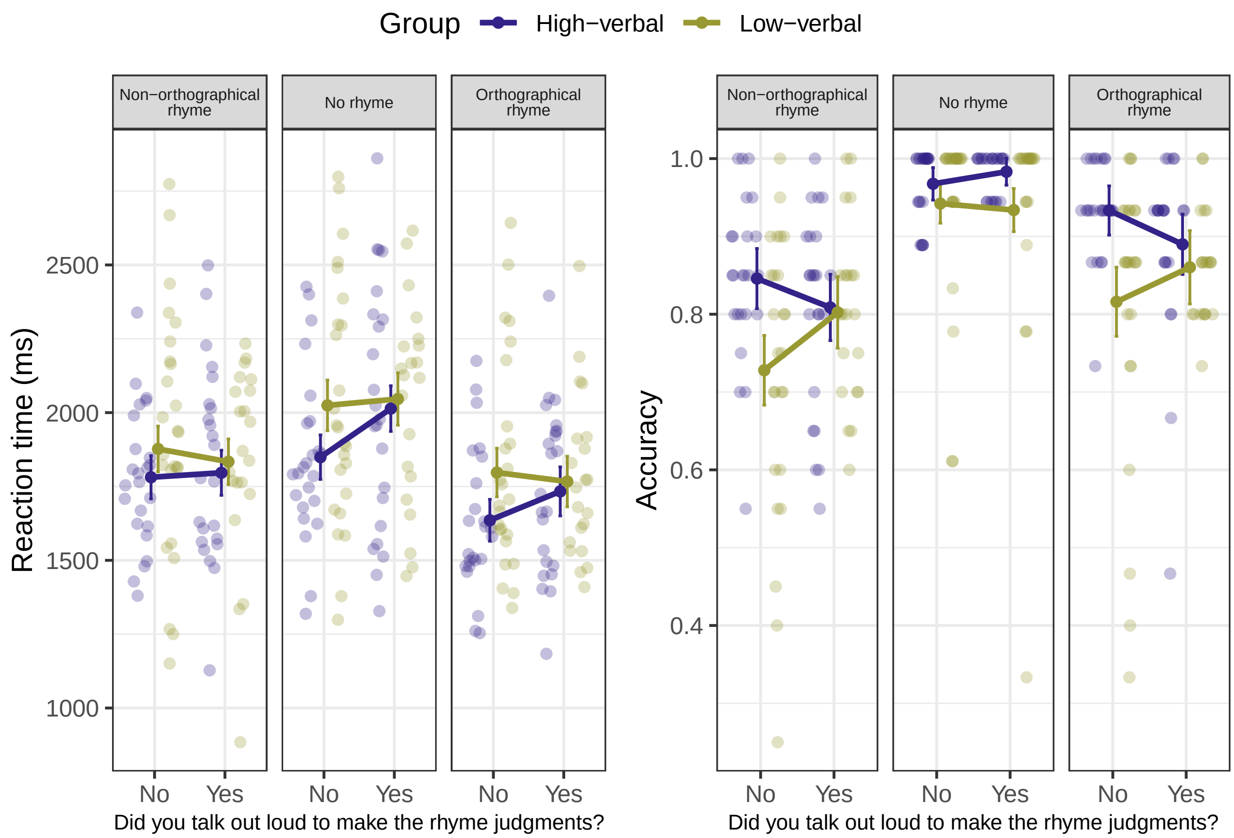

Figure 10. Reaction time and accuracy by whether participants indicated that they had talked out loud to make the rhyme judgments. Error bars indicate 95 % confidence intervals.

between adding and subtracting counted as correct (as long as the arithmetic itself was also correct). We did this to prevent a failure to switch once resulting in the remaining trials counting as incorrect.

3.3.1 Descriptive statistics: Task switching. As can be seen from Table 4 and Figure 11, accuracy was generally quite high in all conditions, and reaction times were comparable across the two groups of participants.

3.3.2 Statistical models: Task switching. We conducted two linear mixed models of verbal score and condition predicting reaction time and accuracy. Both models included random intercepts for condition by participants, and the predictors were both centered. The model predicting accuracy found that condition was a significant predictor with an "increase" in condition being associated with lower accuracy (β = -0.58; SE = 0.10;

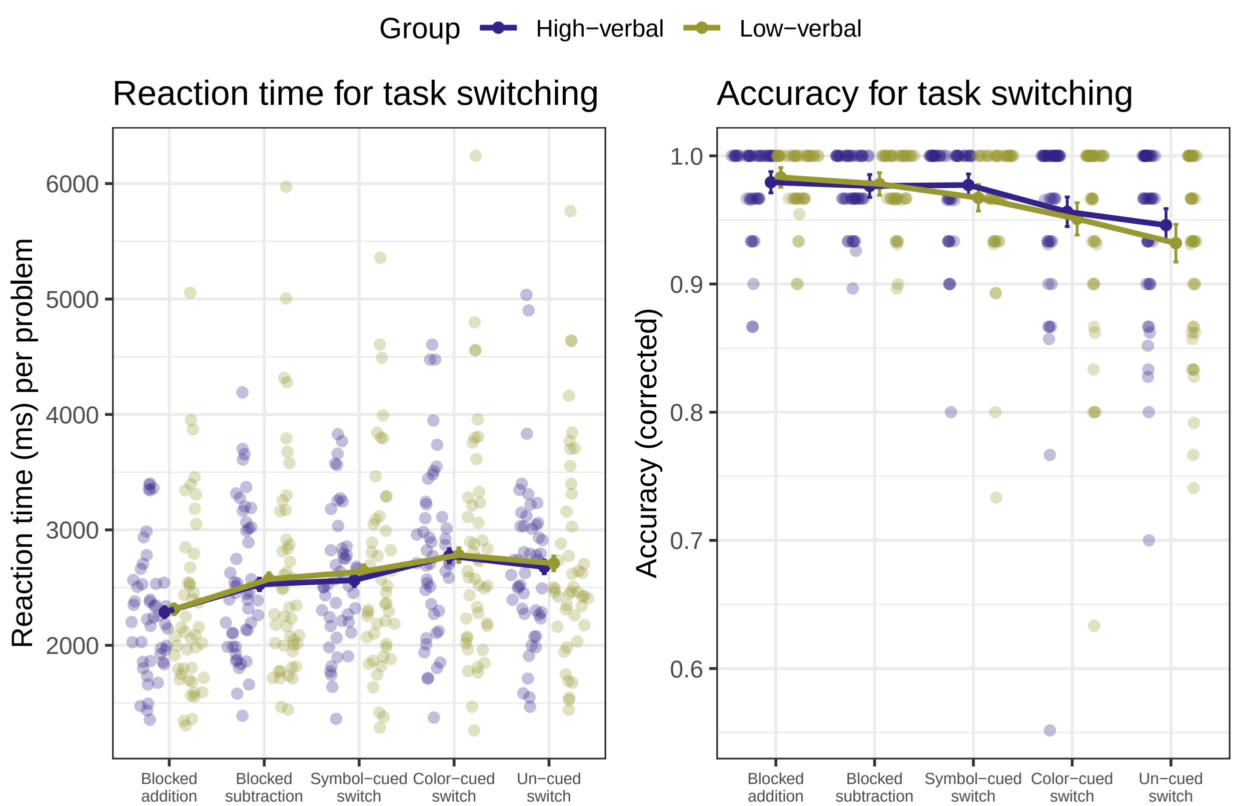

Figure 11. Reaction time and accuracy across conditions in the task switching experiment. Error bars indicate 95 % confidence intervals.

z = -5.50; p < .001). There was no effect of verbal score (p = 0.57). However, there was a marginally significant interaction effect with a higher verbal score diminishing the negative effect of condition difficulty (β = 0.16; SE = 0.08; z = 1.87; p = 0.06). As for log-transformed reaction time, there was a significant effect of condition difficulty with more difficult conditions being associated with slower reaction times (β = 0.07; SE = 0.02; z = 2.99; p = 0.00) but no effect of verbal score and no interaction effects (both p > 0.72).

If we calculate switch cost in a similar way to Emerson and Miyake (2003) (i.e., by averaging problem solution time across simple addition and subtraction and then subtracting the result from average problem solution time in the three switching conditions), we also do not find any effects of verbal score on either reaction time (β = 3.46; SE = 30.99; t = 0.11; p = 0.91) or accuracy (β = -0.01; SE = 0.00; t = -1.21; p = 0.23). However, switch costs differed between the three switching conditions with the uncued switch condition being associated with higher reaction time switch costs (β = -67.29; SE = 22.60; t = -2.98; p = 0.00) and higher accuracy switch costs (β = -0.02; SE = 0.00; t = -4.42; p < .001).

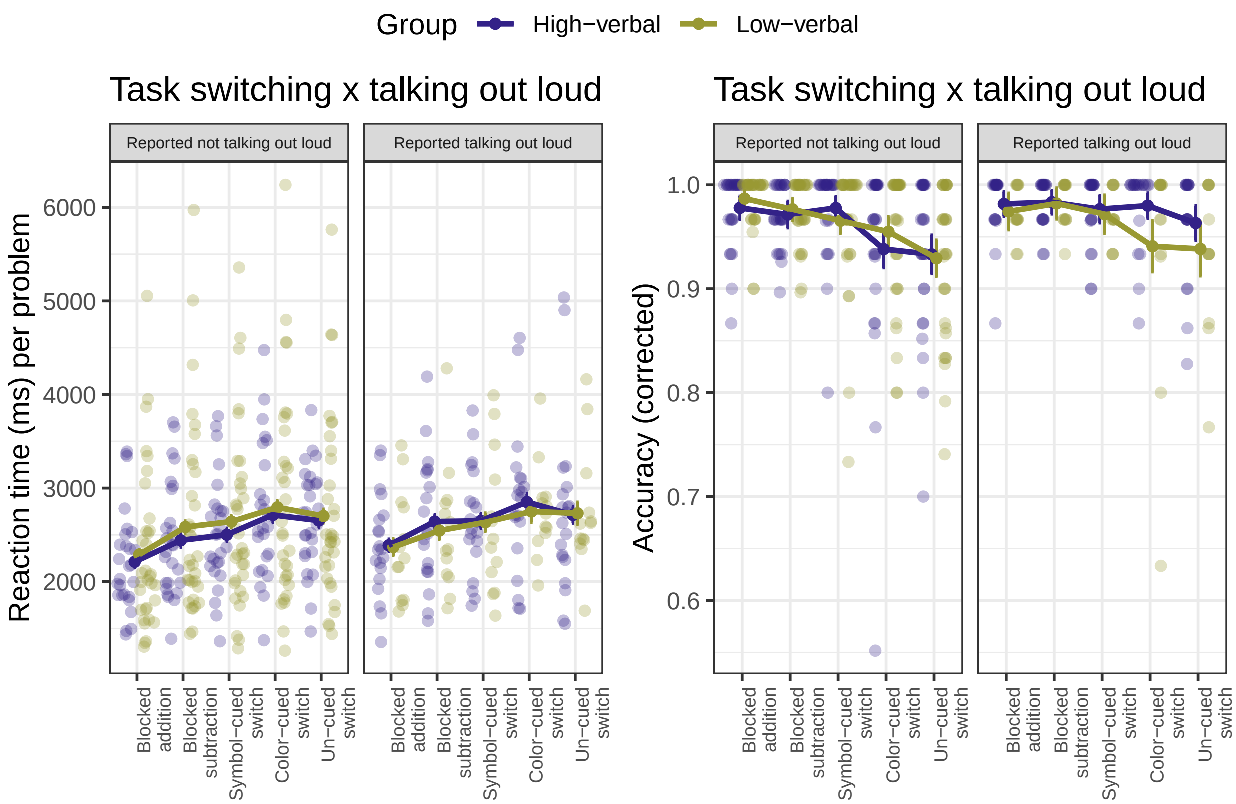

Figure 12. Accuracy in the task switching experiment by whether participants reported talking out loud to remember the correct rule or not.

3.3.3 Strategies: Task switching. We once again examined differences associated with talking out loud, despite the fact that there were no general differences in performance between the two groups. A chi-squared test showed that there was no significant difference between how many high-verbal participants (20 out of 47) and how many low-verbal participants (13 out of 46) reported that they had talked to themselves out loud during the task ($\chi^2(1)$ = 1.50, p = 0.22). There were not any obvious differences between the effects that talking out loud had on these two groups (see accuracy and reaction time Figure 12).

In response to the free answer question in the task switching experiment, several of the high-verbal participants said that they had said the answers out loud to themselves but not the operation ('add', 'subtract'). One visualized a cartoon character wearing red and giving thumbs-up or wearing blue and giving thumbs-down, one used their own thumb to keep track, and one used their fingers to count. Participants from the low-verbal group did not report many specific strategies apart from a few saying the operation or result out loud - one reported that they had tapped their index finger to mean 'add' and their middle finger to mean 'subtract'.

3.4 Same/different judgments

We excluded trials above 5 seconds (0.7 %) and below 200 ms (0.07 %).

3.4.1 Descriptive statistics by group: Same/different judgments. Generally, participants made the correct judgment on 95.53 % of trials. This did not differ between the high-verbal (95.58 %) and the low-verbal group (95.48 %). In subsequent analyses and plots, we only include correct trials. See Figure 13 for reaction times between the high-verbal and low-verbal groups for category judgments ('do these two animals belong to the same category?') or identity judgments ('are these two animals identical?').

3.4.2 Statistical models: Same/different judgments. We conducted a linear mixed model of verbal score and judgment type predicting log-transformed reaction time including random intercepts per judgment type by participant. This model indicated significant main effect of judgment type and a marginally significant effect of verbal score. Identity judgments were faster than category judgments (β = -0.10; SE = 0.01; t = -11.38; p < .001), and a higher verbal score was marginally associated with faster reaction times (β = -0.04; SE = 0.02; t = -1.80; p = 0.08).

The key test for this experiment was whether the two groups behaved differently when giving correct 'DIFFERENT' responses on identity trials when the two images belonged to the same category. That is, we expected high-verbal participants to be more susceptible to

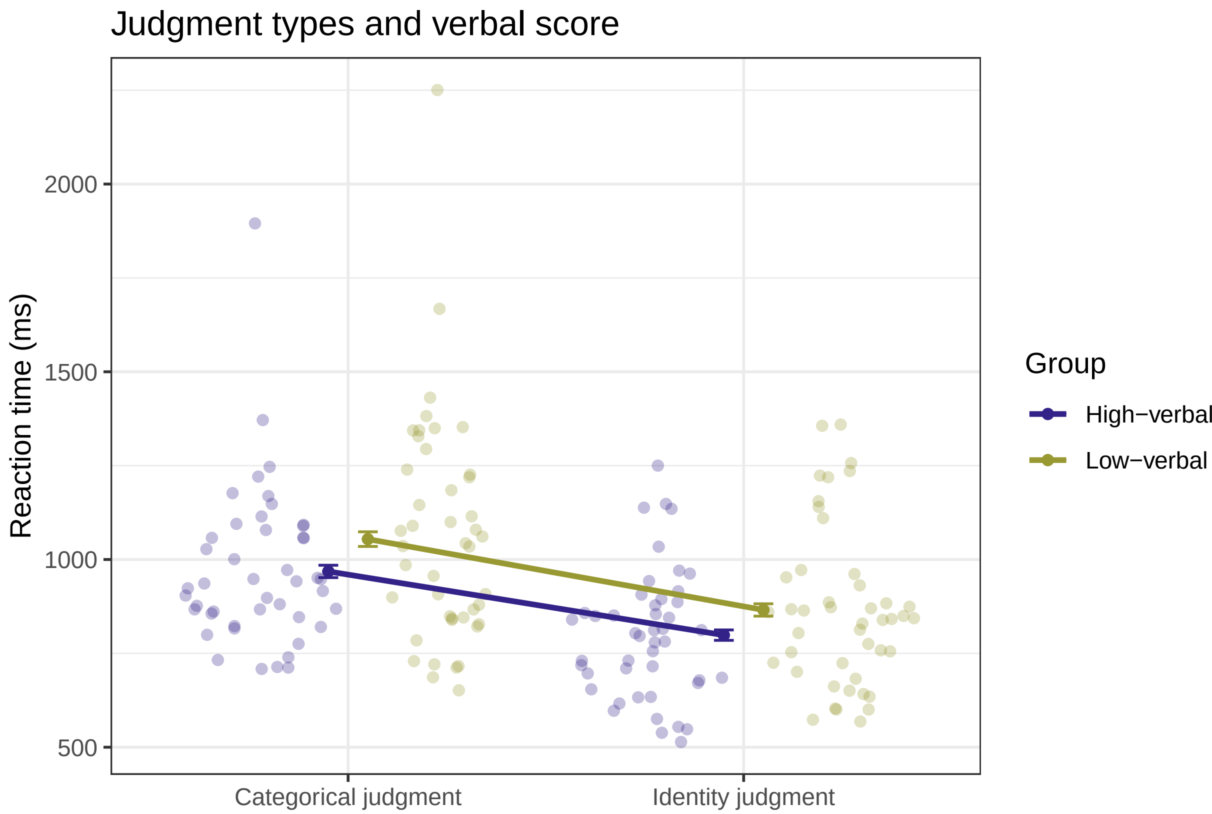

Figure 13. Reaction time in response to category or identity judgments. Error bars indicate 95 % confidence intervals.

interference from a same-category distractor. See Figure 14.

A linear mixed model of log-transformed reaction time with verbal score and category membership of the distractor as predictors, including random intercepts per category membership by participant, provided evidence that high-verbal participants were not particularly affected by the within-category interference (interaction effect: p = 0.95). However, there was a significant main effect of category membership of the distractor with within-category distractors being associated with slower reaction times (β = -0.04; SE = 0.01; t = -7.71; p < .001).

3.4.3 Additional analyses: Same/different judgments. We also checked whether the kind of animal made a difference on a within-category distractor trial. See Figure 15.

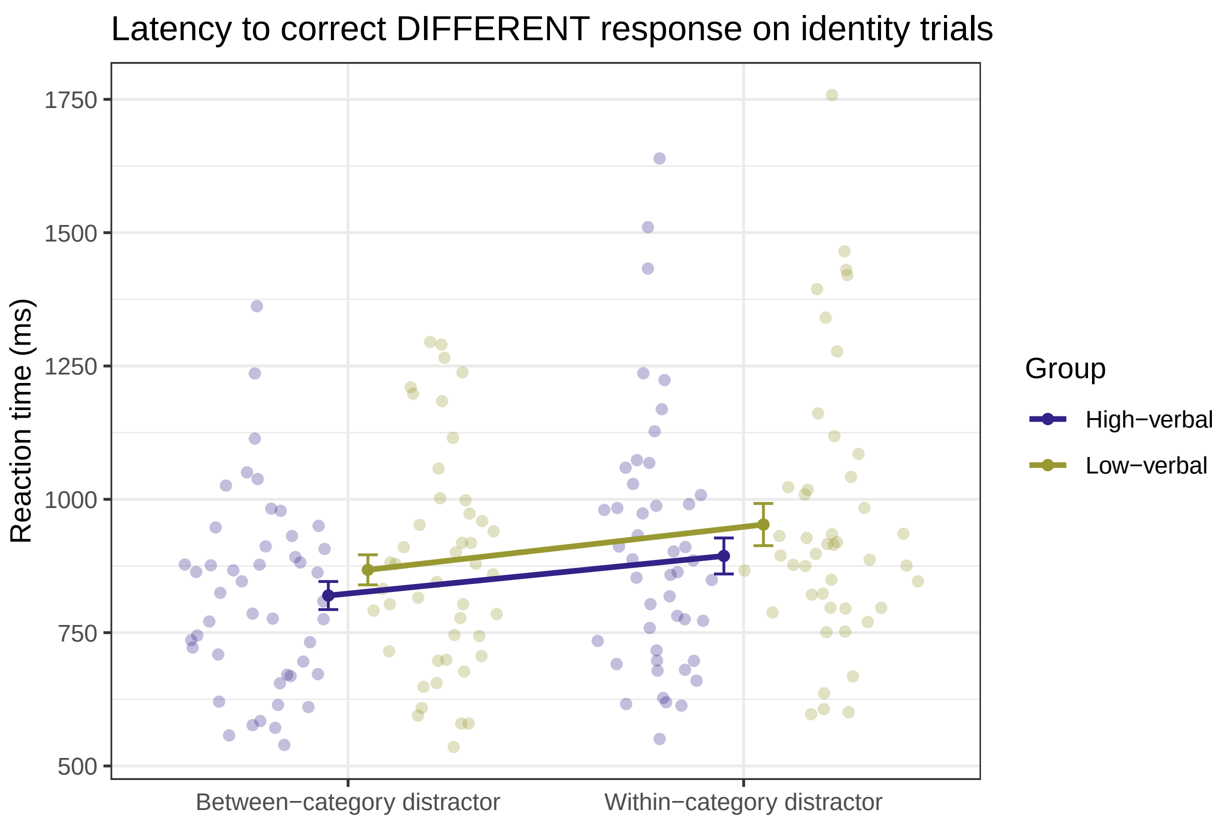

Figure 14. Reaction time on identity trials where the correct response was ‘DIFFERENT’ either because the two silhouettes were from different categories or different images from the same category. Error bars indicate 95 % confidence intervals.

A linear mixed model of log-transformed reaction times with verbal score and animal pair (dog-dog or cat-cat) as predictors, including random intercepts per animal pair by participant, provided evidence that dog-dog trials were faster than cat-cat trials (β = -0.03; SE = 0.01; t = -3.85; p < .001). However, this effect of animal pair was less strong when verbal score was higher as indicated by a significant interaction effect between verbal score and animal pair (β = 0.03; SE = 0.01; t = 3.11; p = 0.00).

3.4.4 Strategies: Same/different judgments. In this experiment, most participants said that they had no particular strategy. However, eight of the high-verbal participants and one of the low-verbal participants explicitly mentioned something to do with verbalizing the problems (e.g. ‘In my head I said “same” or “different” before I pressed

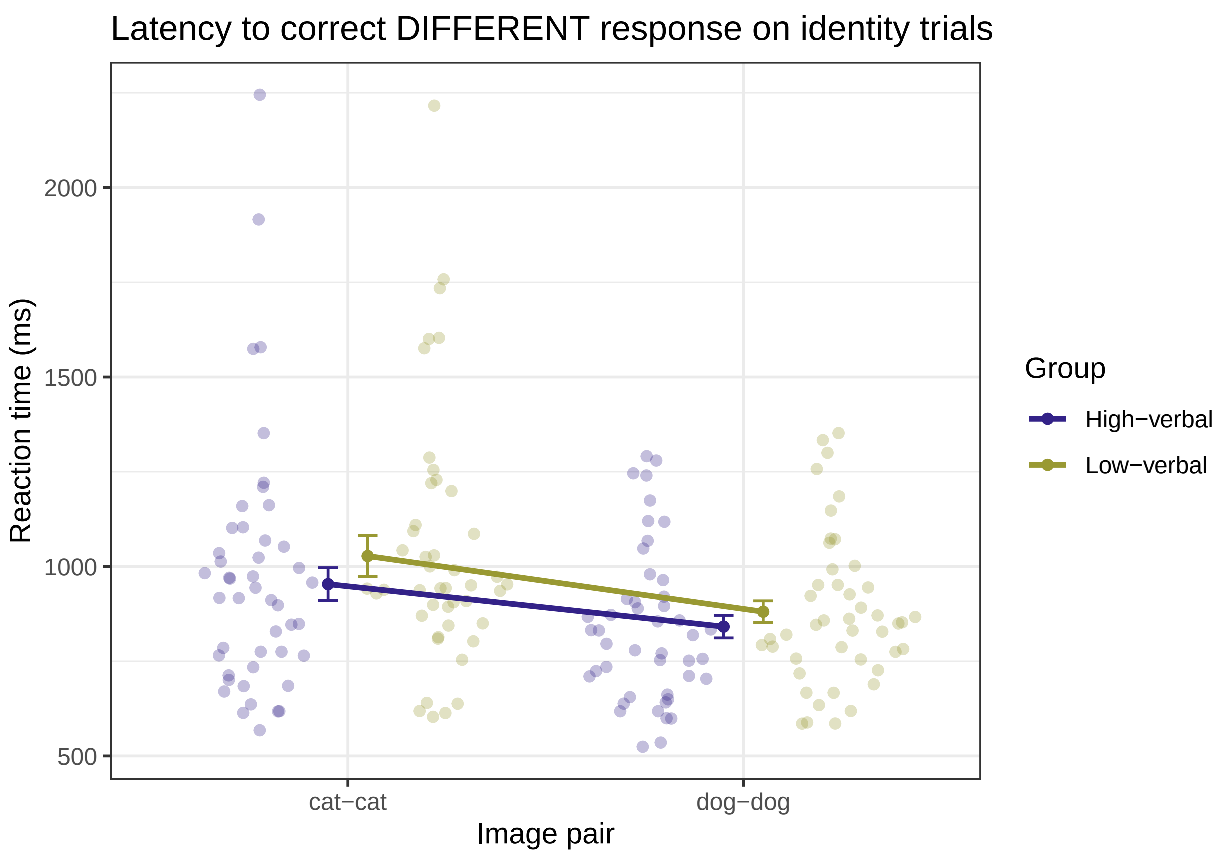

Figure 15. Reaction time in response to cat-cat or dog-dog judgments on identity trials when the correct response was ‘DIFFERENT’ (i.e., one cat and another cat or one dog and another dog). Error bars indicate 95 % confidence intervals.

the arrow key.’)

3.5 Intertask correlations

We were interested in how performance on the different tasks correlated with each other and whether these correlations were different for the two groups.

3.5.1 Overall intertask correlations. See Figure 16. Generally, different performance measures correlated within the same experiment. Interestingly, reaction times on rhyming were negatively correlated with verbal working memory score suggesting some working memory involvement in the rhyming task. Accuracy on uncued switch trials in the task switching experiment was also positively correlated with accuracy on verbal working memory and rhyming.

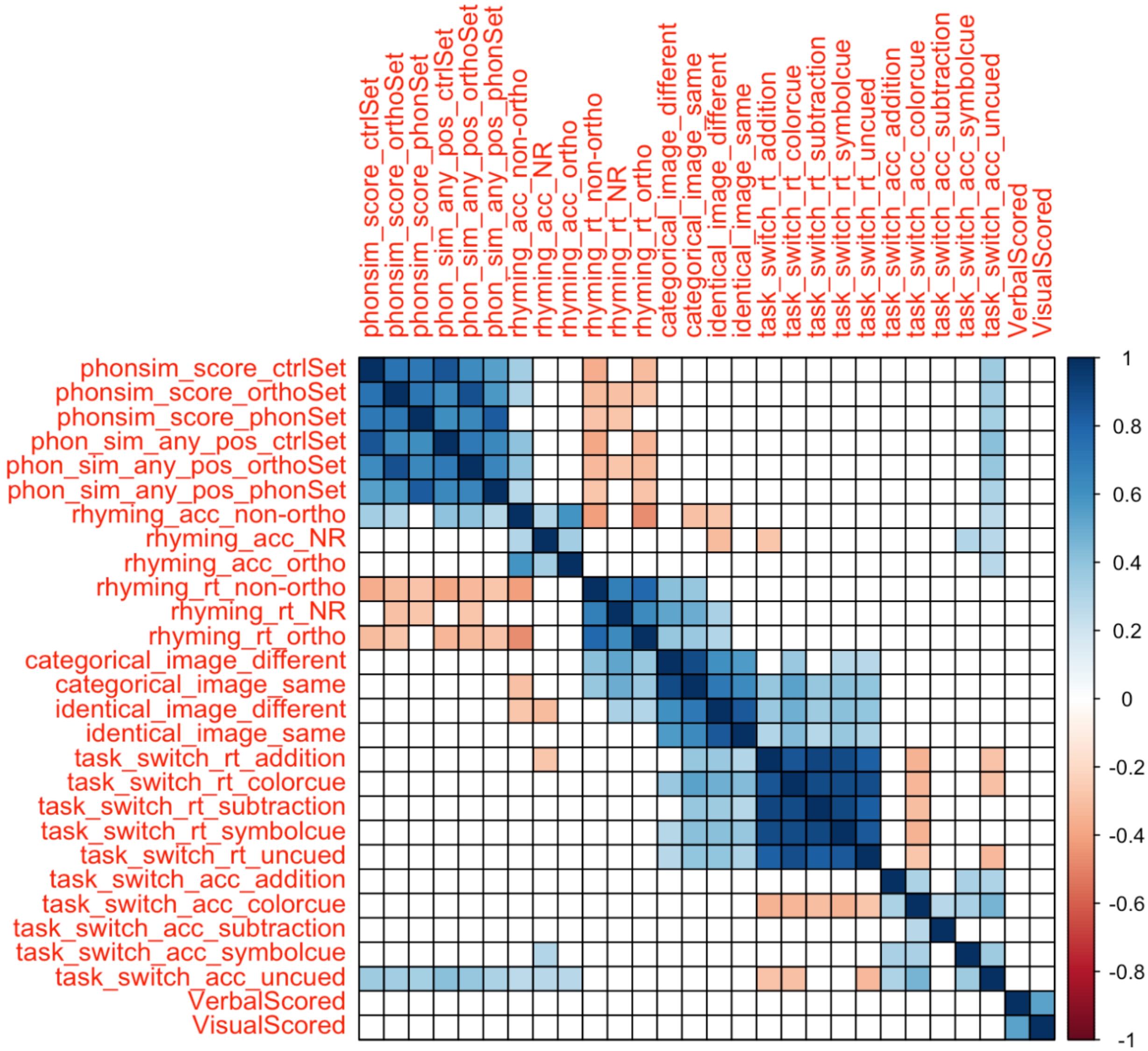

Figure 16. Intertask correlations in the total sample of high-verbal and low-verbal participants. Colored squares represent significant correlations at $p < .01$.

3.5.2 Intertask correlations for the *high-verbal group.*

See Figure 17. For high-verbal participants, reaction times on rhyming, category judgments, and task switching were positively correlated, suggesting that these relied on similar mechanisms for this group.

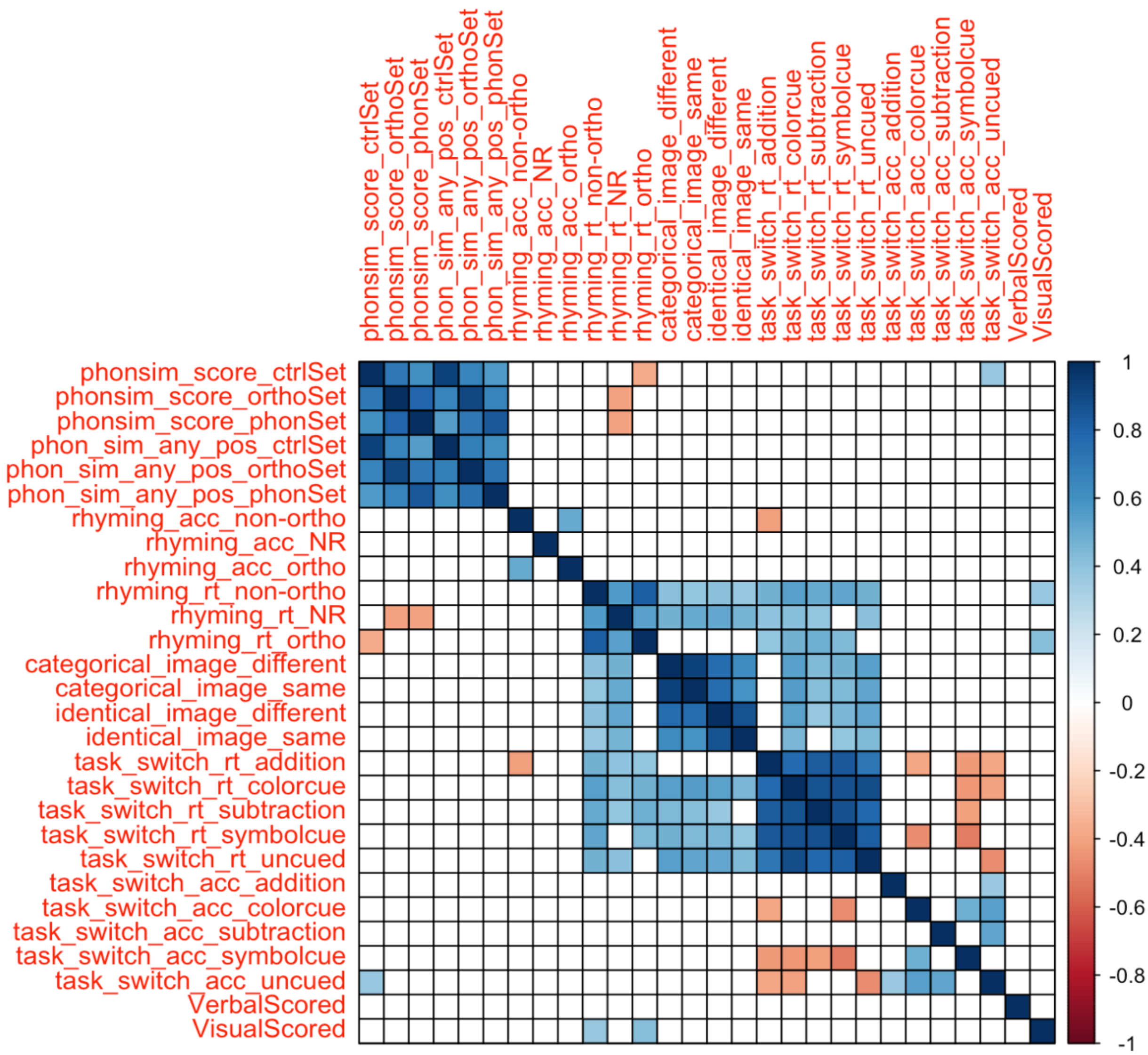

Figure 17. Intertask correlations within the high-verbal group. Colored squares represent significant correlations at p < .01.

3.5.3 Intertask correlations for the *low-verbal group*.

See Figure 18. For low-verbal participants, there was no such widespread correlation between rhyming, category judgments, and task switching reaction times. However, they showed a positive correlation between accuracy on uncued switch trials and (position-indifferent) verbal working memory and rhyming accuracy.

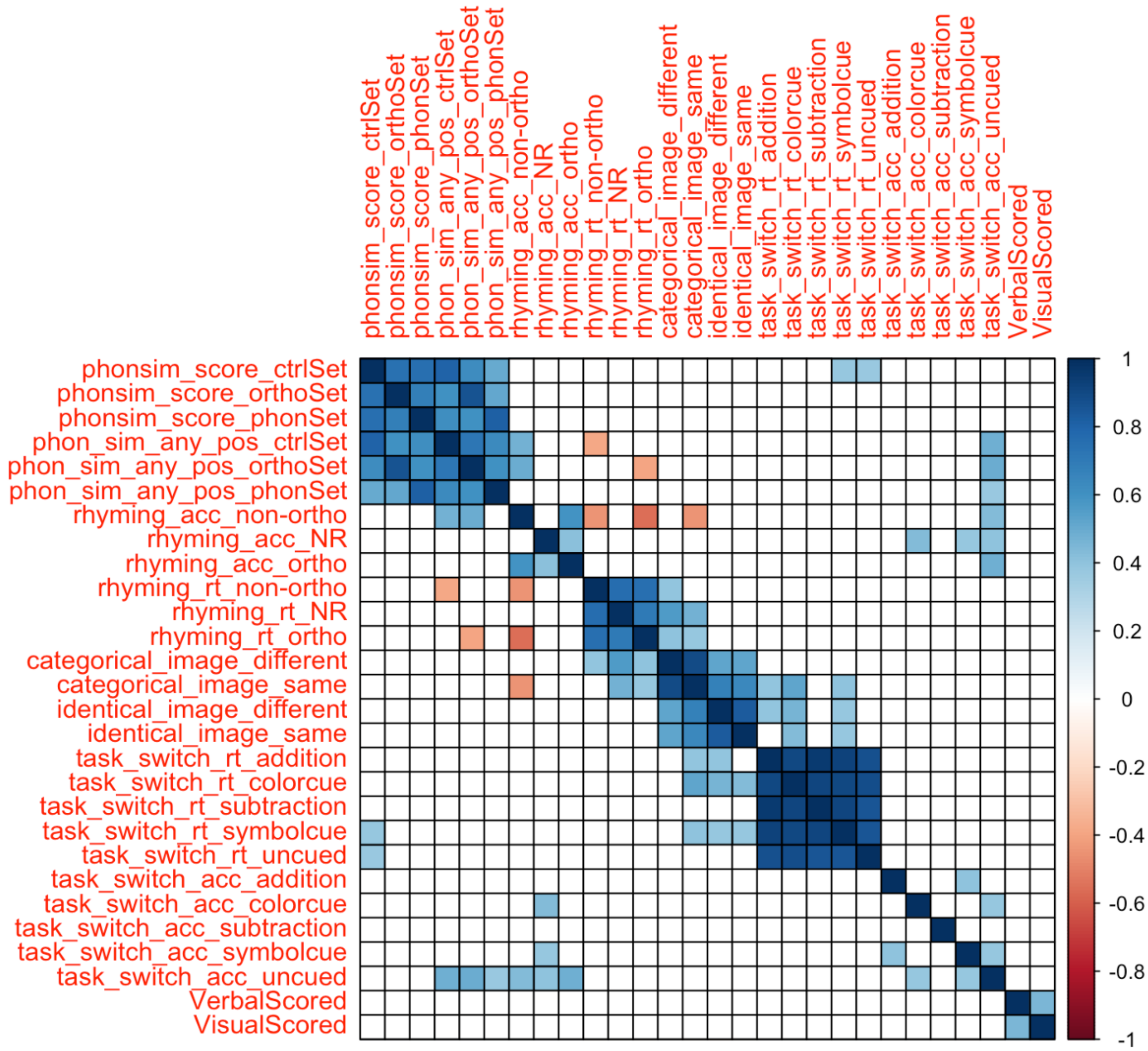

Figure 18. Intertask correlations within the low-verbal group. Colored squares represent significant correlations at $p < .01$.

3.5.4 Difference between task correlations in *low-verbal group* and *high-verbal group*.

See Figure 19. Note that colored squares do not have any relationship with statistical significance on this plot. Blue simply means that when the correlation coefficient for the high-verbal group was subtracted from the correlation coefficient for the low-verbal group, the result was positive. Similarly, red-orange means that when the correlation coefficient for the high-verbal group was subtracted from the correlation coefficient for the low-verbal group, the result was negative.

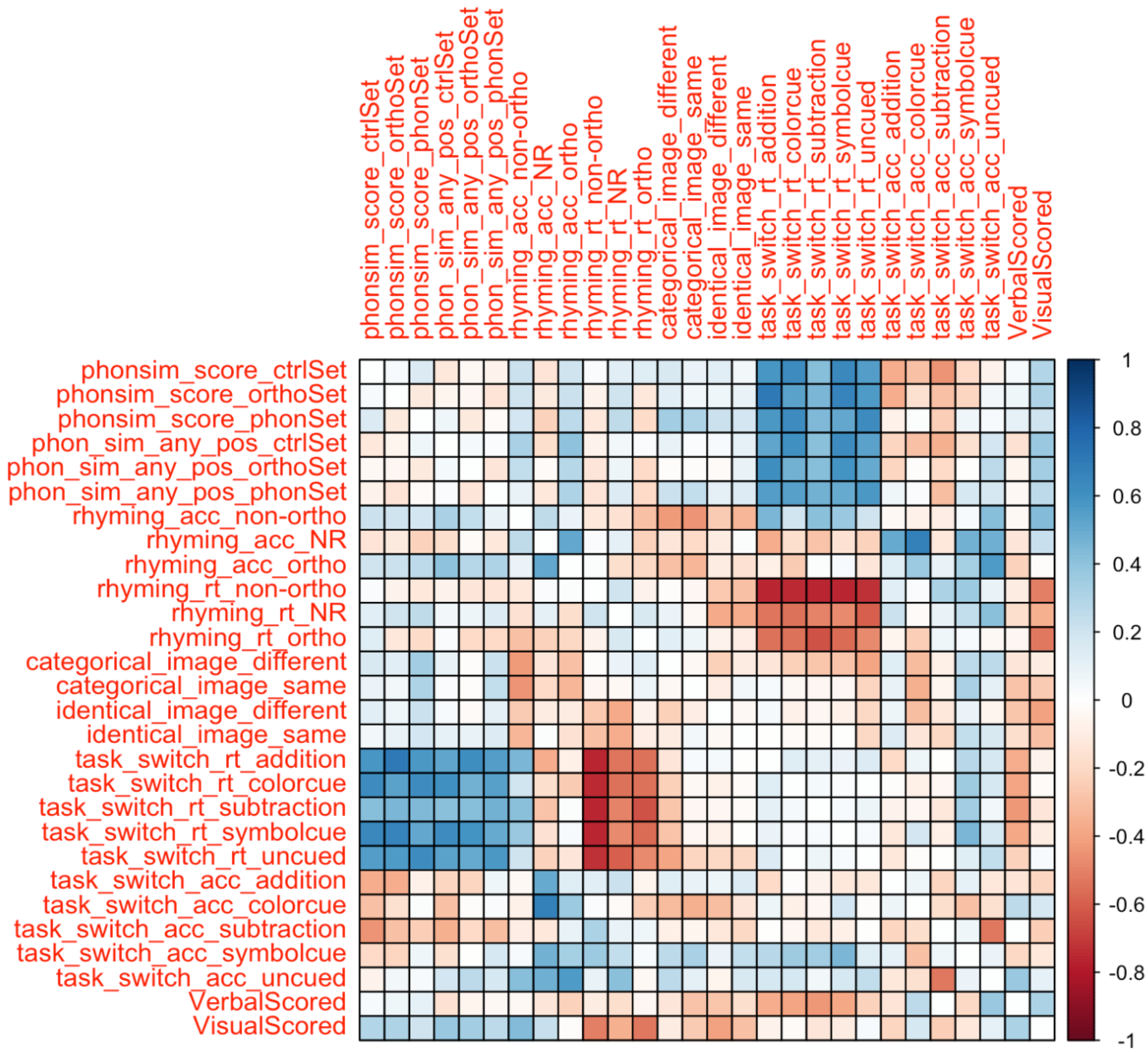

Figure 19. Intertask correlations in the high-verbal group subtracted from intertask correlations in the low-verbal group.

3.6 Summary of behavioral findings

In the phonological similarity experiment, the high-verbal group performed better at both position-specific recall and position-indifferent recall. The difference between the two groups was diminished for the orthographically similar set when considering position-indifferent recall. Interestingly, the differences between the two groups was diminished if participants reported talking out loud to remember the words.

In the rhyming experiment, high-verbal participants performed significantly better than low-verbal participants. Once again, the differences in both accuracy and reaction time between the two groups was eliminated when participants reported naming the pictures out loud.

Participants who scored lower on verbal representations were slower at making category-based judgments in the same-different experiment. We expected high-verbal participants to show a more marked effect of being distracted by category membership on identity judgment trials (e.g. being slower at correctly responding ‘different’ to two pictures of different cats), but this did not appear to be the case.

There were no notable differences between the two groups in the task switching experiment.

3.7 Questionnaire measures

For some strange reason, we do not have questionnaire data from A3KVKK1XLBTSN3. We will retain their data from the four behavioral experiments and here report questionnaire data from 47 high-verbal and 45 low-verbal participants.

For most of our custom questions, there were notable differences in how participants from the two groups responded (see Figures 20 and 21). The questions with the clearest differences concerned rehearsing and revising conversations where the high-verbal group reported doing so much more often than the low-verbal group did.

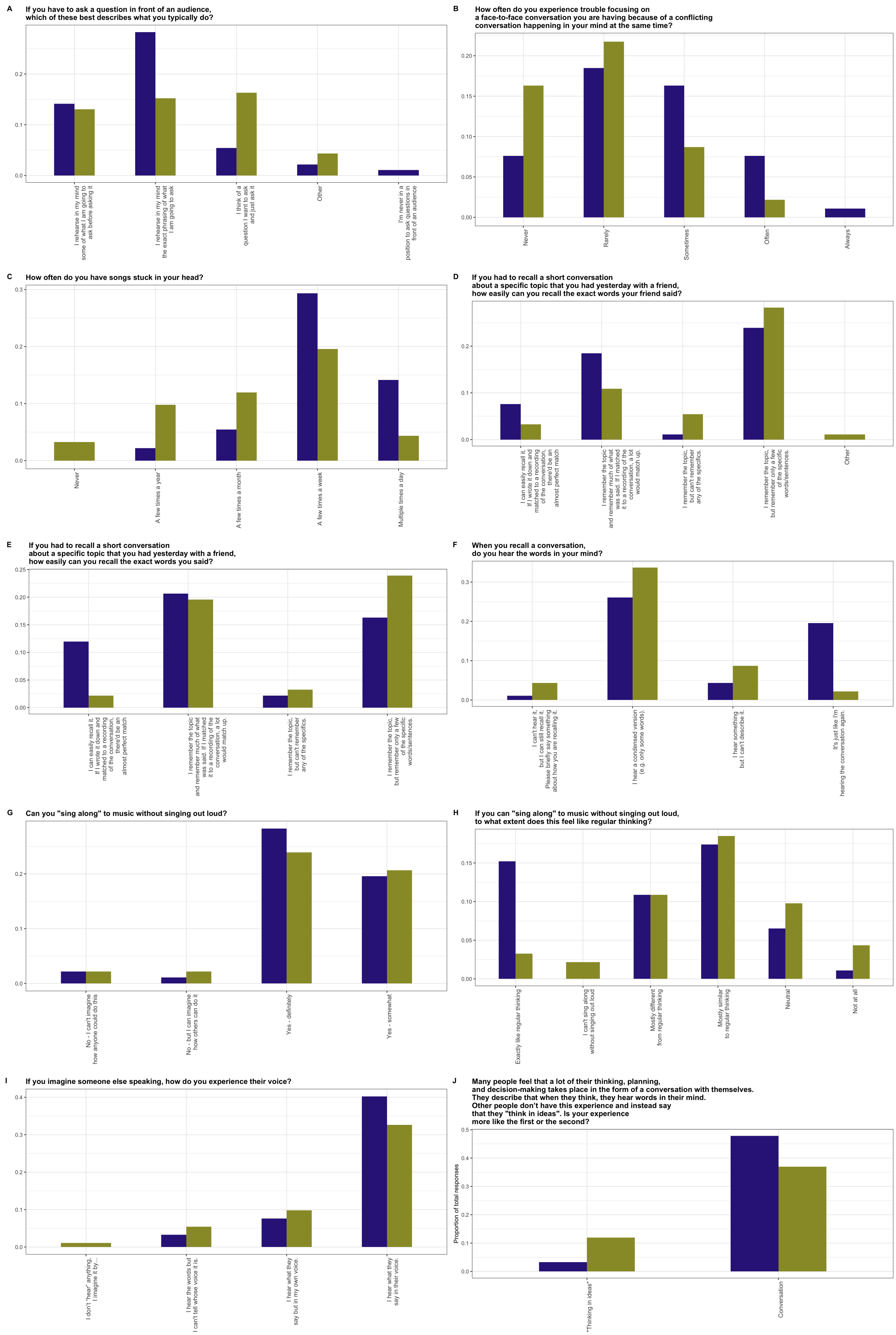

Figure 20. Grouped bar plots of our custom questions. Dark blue represents the high-verbal group and dark yellow represents the low-verbal group.

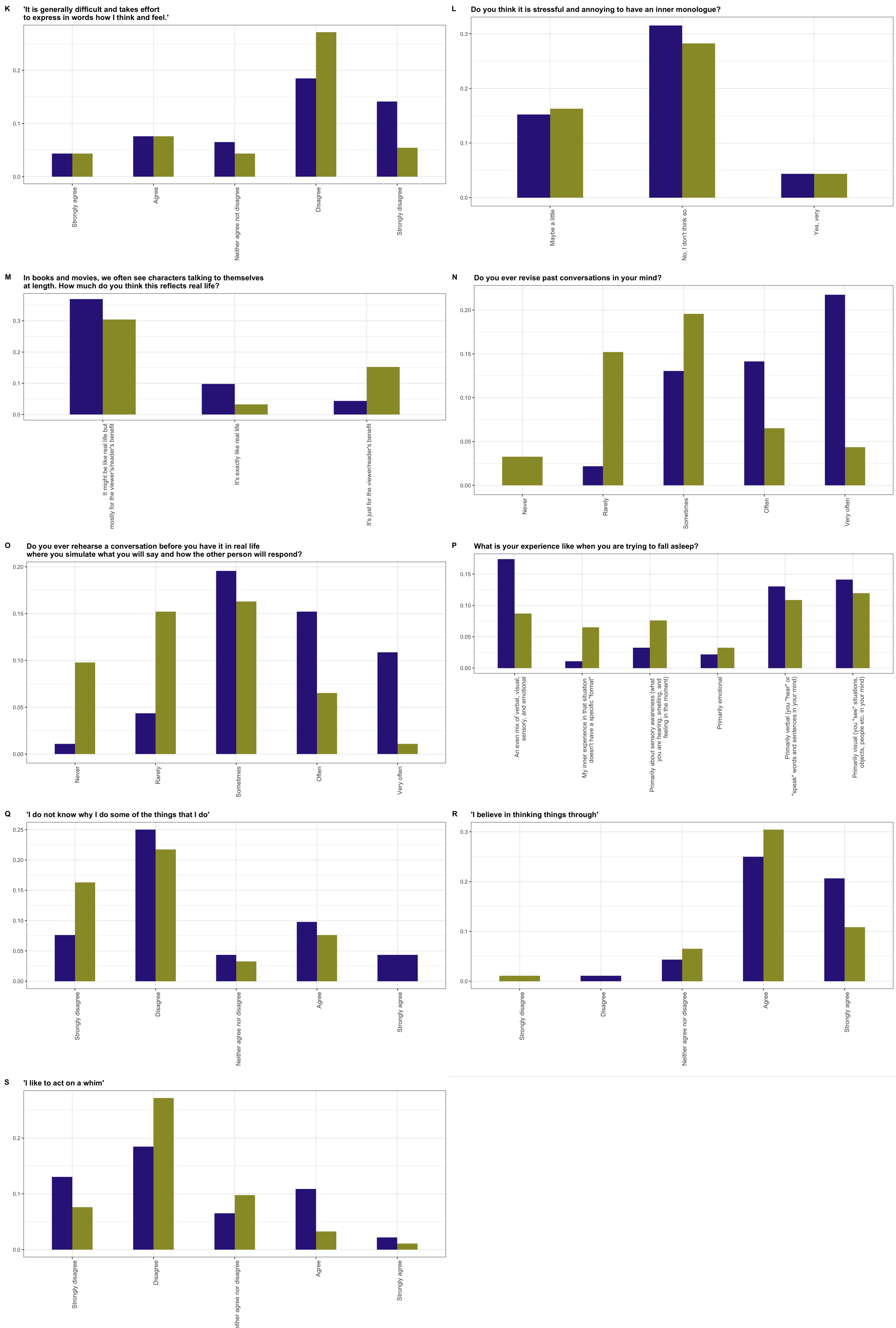

Figure 21. Grouped bar plots of our custom questions. Dark blue represents the high-verbal group and dark yellow represents the low-verbal group.

4 Discussion

In this exploratory study, we found significant behavioral differences between participants who experience a lot of verbal representations and participants who do not. High-verbal participants were more accurate at making rhyme judgments from images, faster at making both category and identity judgments for simple visual stimuli, and showed better verbal working memory performance. The only experiment where we found no performance differences between the two groups was the task switching experiment. Interestingly, in both the rhyming experiment and the verbal working memory experiment, performance differences between the two groups disappeared when participants reported talking out loud to solve the problems. In terms of our questionnaire, the differences were as expected: High-verbal participants were more likely to say that they rehearse or revise conversations, and that their experience of thinking about or recalling conversations is much like hearing every word. They also reported having songs stuck in their head more often and were less likely to say that the way inner speech is represented in media is for the reader's/viewer's benefit.

Having found evidence that a diminished experience of inner speech has behavioral consequences, we would like to propose a name for the phenomenon: *anendophasia*. We came up with this in collaboration with people on Reddit and in real life who experience no inner speech. Many of them objected to words implying silent or quiet minds as they experience their minds as quite busy. Similarly, they did not find words implying lack of mental language appropriate as they experience words/concepts but not speech. We believe anendophasia is suitable because it includes components that people are likely to already be familiar with ('-phasia' as something to do with speech) and targets specifically speech and not words per se. Furthermore, the term endophasia already exists as a term for inner speech (Bergounioux, 2001; Loevenbruck et al., 2018). We do not believe a lack of inner speech could simply be subsumed under the umbrella term 'aphantasia' for lack of mental imagery (Monzel, Mitchell, Macpherson, Pearson, & Zeman, 2022) because inner speech is both auditory and articulatory in nature (whether it is better termed 'inner hearing' or 'inner speaking' is also subject to debate) and because the linguistic properties of inner speech are not reducible to phonological properties. For these reasons, we also do not believe the previously proposed term *anauralia* is appropriate (Hinwar & Lambert, 2021).

4.1 What have we learned about people with anendophasia?

Most importantly, it appears that people are right when they report that their experience rarely takes a verbal format. This is not self-evident as Descriptive Experience Sampling has sometimes found that people can be quite mistaken about the format of their inner experience (Heavey & Hurlburt, 2008; Russell T. Hurlburt, 2011; Russell T. Hurlburt, Heavey, & Kelsey, 2013). On our tasks, participants with anendophasia seemed less able to use internal verbalization to increase performance. This is especially interesting as the questions that are related to the verbal factor on the Internal Representations Questionnaire (Roebuck & Lupyan, 2020) and which we used for participant selection are about the format of spontaneous thought (e.g., 'I think about problems in my mind in the form of a conversation with myself' and 'If I am walking somewhere by myself, I often have a silent conversation with myself'). There is some evidence that spontaneously occurring inner speech and experiment-elicited inner speech are not necessarily comparable and have different neural subtrates (Russell T. Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016) which makes it remarkable that our participants' reports of spontaneous inner speech seem related to their ability to use internal verbalization and verbal working memory. It is also interesting that performance was in many cases related to verbal score as a continuous factor which indicates that anendophasia is not an all-or-nothing phenomenon, much like aphantasia is not (Dance et al., 2022).

Regarding the parallels with aphantasia, it is important to note that the analogy can only take us so far. Given the findings from the present study, it seems unlikely that people with anendophasia are completely unable to verbalize internally like some people with aphantasia are completely unable to visualize. If you ask somebody with anendophasia to read something out loud in their head, for example, they can do it. What seems instead to be the case is that they do not use or only rarely use inner speech spontaneously in everyday life to plan, solve problems, and rehearse conversations (Perrone-Bertolotti et al., 2014). A question remains of how thinking is experienced if not verbally. As discussed in the Introduction, there is anecdotal evidence that some people with anendophasia have very visual experience and others “think in ideas”. This largely corresponds to what we found when we asked participants whether their thinking was more like a conversation or more like “thinking in ideas”. One participant described their thinking as ‘I really do think in concepts rather than forming words in my head’ and another reported ‘I visualize what I am trying to do or plan and act accordingly’. We are not much closer to understanding what “thinking in ideas” means but note that it is probably related to the unsymbolized thinking of Hurlburt and colleagues.

We did find evidence that using other strategies that internal verbalization could diminish the performance differences between our two groups. This was clearest when we examined whether participants reported talking out loud to solve the problems or not. In both the verbal working memory experiment and in the rhyme judgment experiment, performance differences disappeared when participants reported talking out loud. This suggests that participants without anendophasia were already using verbalization strategies internally. One particularly interesting example comes from orthographically similar words in the verbal working memory experiment (‘rough’, ‘cough’, ‘through’, ‘dough’, ‘bough’). Many participants with anendophasia reported a strategy of remembering just the first letters of the words once they were familiar with the set, thus reducing the load on verbal working memory. This is likely to be the reason why there was no difference in performance between the two groups for this word set. Another interesting case is the finding that the two groups did not differ in either reaction time or performance on the task switching experiment. This suggests that while the inner voice can be used as a behavioral self-cue, other and equally effective strategies may be available.

4.2 Implications for the relationship between language and cognition

In our view, the most interesting question that the present paper can help address is this: If there are indeed people who have no inner voice and do not use language to think, and if these people behave more or less like everyone else (hence the masked individual differences), does this mean that the connection between language and thought could in fact be quite weak? We saw that participants with anendophasia performed worse on tasks specifically requiring internal verbalization (verbal working memory and rhyme judgments) and were also slower on both category and identity judgments. However, they did not differ from our high-verbal control group in performance on a task switching paradigm – this is surprising because verbal interference studies have previously suggested that people use their inner voice to cue themselves on the relevant task rule (plus or minus). Our interpretation of our findings is that inner speech can be recruited for self-cuing but that self-cuing can be as effectively achieved through other means as well.

Taking the findings from our experiments together, it seems unlikely that inner speech plays a constitutive role in cognition, although we allow that this may still be the case for some kinds of cognition (e.g., metacognition and self-reflection, see Morin (2018) and Bermúdez (2007)). Inner speech is instead plausibly recruited as a tool by some people for some tasks which is a much weaker connection than has previously been suggested. When inner speech is recruited for tasks, it provides format that supports linear order (as seen in our verbal working memory experiment) and more readily available categorization (as seen in our same-different experiment). Our study is merely a preliminary exploration, and there are many other kinds of thinking that should be tested as a next step. Candidates include abstraction, relational thinking, metacognition, and social cognition, all of which have been hypothesized to rely on language to some extent.

4.3 Limitations of the present study

It may be the case that the performance differences we found were driven by some other factor than how verbally represented participants' thoughts were. For example, high-verbal participants could be generally more motivated or conscientious than low-verbal participants. We believe that this explanation is unlikely, however, since we saw examples of specific conditions where there were no differences between the two groups (e.g., responses to the no-rhyme pairs in the rhyme judgment experiment, responses to the orthographically similar set in the verbal working memory experiment, and all conditions in the task switching experiment).

4.4 Future directions

The present work is exploratory and presents a few examples of how we might empirically test the consequences of having or not having an inner voice. Some of the most interesting avenues forward concern alternative strategies as well as how a lack of inner speech might influence other things than behavior such as mental health and social cognition.

Just as in aphantasia, it could be the case that individual differences in inner speech remain largely undiscovered because people use alternative but equally efficient strategies for solving problems. We saw this for example in the aphantasia study where participants with aphantasia performed equally well but had a different response profile (Keogh et al., 2021). We see some indications in our present study as well with the intertask correlations being different in the low- and high-verbal groups. This should be explored in future studies – the optimal way would be to design an experiment where different strategies would show different behavioral profiles but this is quite difficult.

Many areas of research and therapy (most notably Cognitive Behavior Therapy and many mindfulness practices) rely heavily on inner speech and are difficult to tailor to people who find this unnatural. A natural first step would be to ask people how much they spontaneously use their inner voice in everyday life and take this into account when designing interventions.

The largest difference between the two groups in the questionnaire was that high-verbal participants spent more time rehearsing conversations which makes us wonder what kind of consequences this might have. Would we expect high- verbal people to be somehow “better” at conversations? Or maybe worse because they over-rehearse? It seems that inner speech is very linked to social interactions so in future studies we would like to assess social cognitive abilities in populations with and without habitual inner speech. This could for example be with an adult version of the Faux Pas test (e.g., Baron-Cohen, ORiordan, Stone, Jones, & Plaisted, 1999; Thiébaut et al., 2015).

5 Conclusion

In four exploratory behavioral experiments, we found significant differences between participants whose inner experience is verbal to a large extent and participants whose inner experience is verbal to a much lesser extent. We propose a term for the experience of the latter group: anendophasia. Participants with anendophasia were worse at making rhyme judgments in response to images, remembering words in order, and slower at making both category and identity judgments of simple visual stimuli. They were not, however, either slower or less accurate at a task switching experiment requiring them to switch between addition and subtraction problems. Taken together, our experiments suggest that there are real behavioral consequences of experiencing less or more inner speech, and that these differences may usually be masked because people with anendophasia use alternative strategies. Our findings have important implications for theories about the relationship between inner speech and cognition.

6 References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931.

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire–revised (VISQ-r): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48–58.

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown.* Retrieved from https://github.com/crsh/papaja

Baddeley, A. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362–365.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641.

Baron-Cohen, S., ORiordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). *Journal of Autism and Developmental Disorders*, *29*(5), 407–418. https://doi.org/10.1023/a:1023035012436

Barth, M. (2022). *tinylabels: Lightweight variable labels.* Retrieved from https://cran.r-project.org/package=tinylabels

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Bergounioux, G. (2001). Endophasie et linguistique [décomptes, quotes et squelette]. *Langue Francaise*, *132*, 106–124.

Bermúdez, J. L. (2007). *Thinking without words.* Oxford University Press.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674.

Carruthers, P. (2009). Mindreading underlies metacognition. *Behavioral and Brain Sciences*, *32*(2), 164–182.

Chella, A., & Pipitone, A. (2020). A cognitive architecture for inner speech. *Cognitive Systems Research*, *59*, 287–292.

Clark, A. (1998). *Language and thought: Interdisciplinary themes* (P. Carruthers & J. Boucher, Eds.). Cambridge University Press.

Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science*, *2*(4), 631–642.

Dance, C., Ipser, A., & Simner, J. (2022). The prevalence of aphantasia (imagery weakness) in the general population. *Consciousness and Cognition*, *97*, 103243.

Dawes, A. J., Keogh, R., Andrillon, T., & Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports*, *10*(1), 1–10.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table

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

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*(1), 148–168.

Faw, B. (2009). Conflicting intuitions may be based on differing abilities: Evidence from mental imaging research. *Journal of Consciousness Studies*, *16*(4), 45–68.

Felton, J. (2020). People with no internal monologue explain what it's like in their head.

IFLScience. Retrieved from https://www.iflscience.com/people-with-no-internal-monologue-explain-what-its-like-in-their-head-57739

Frankish, K. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente, Eds.). Oxford University Press.

Galton, F. (1880). Statistics of mental imagery. *Mind*, *5*(19), 301–318.

Gauker, C. (2011). *Words and images: An essay on the origin of ideas.* Oxford University Press, Oxford.

Geva, S. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente, Eds.). Oxford University Press.

Geva, S., Bennett, S., Warburton, E. A., & Patterson, K. (2011). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology*, *25*(3), 323–343.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98.

Harrell Jr, F. E., Charles Dupont, with contributions from, & others., many. (2021). *Hmisc: Harrell miscellaneous.* Retrieved from https://CRAN.R-project.org/package=Hmisc

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, *17*(3), 798–810.

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Hinwar, R. P., & Lambert, A. J. (2021). Anauralia: The silent mind and its association with aphantasia. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.744213

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., et al.others.

(2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*, *104*(10), 4234–4239.

Hurlburt, Russell T. (2011). *Investigating pristine inner experience.* Cambridge University Press. https://doi.org/10.1017/cbo9780511842627

Hurlburt, Russell T., & Akhter, S. A. (2006). The descriptive experience sampling method. *Phenomenology and the Cognitive Sciences*, *5*(3), 271–301.

Hurlburt, Russell T., & Akhter, S. A. (2008). Unsymbolized thinking. *Consciousness and Cognition*, *17*(4), 1364–1374.

Hurlburt, Russell T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the ecological validity of thinking on demand: Neural correlates of elicited vs. Spontaneously occurring inner speech. *PLoS One*, *11*(2), e0147932.

Hurlburt, Russell T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. *Cortex*, *105*, 61–73.

Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots.* Retrieved from https://CRAN.R-project.org/package=ggpubr

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60.

Keogh, R., Wicken, M., & Pearson, J. (2021). Visual working memory in aphantasia: Retained accuracy and capacity with a different strategy. *Cortex*, *143*, 237–253.

Kothe, E., Callegher, C. Z., Gambarota, F., Linkersdörfer, J., & Ling, M. (2021). *Trackdown: Collaborative writing and editing of r markdown (or sweave) documents in google drive.* https://doi.org/10.5281/zenodo.5167320

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

https://doi.org/10.18637/jss.v082.i13

@KylePlantEmoji. (2020). Fun fact: Some people have an internal narrative and some don't. Twitter. Retrieved from https://twitter.com/KylePlantEmoji/status/1221713792913965061?s=20&t=JMap2sXmuh-XA7h5CDxZ5A

Langland-Hassan, P. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente, Eds.). Oxford University Press.

Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, *6*, 528.

Loevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., … Perrone-Bertolotti, M. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente, Eds.). Oxford University Press.

Lurito, J. T., Kareken, D. A., Lowe, M. J., Chen, S. H. A., & Mathews, V. P. (2000). Comparison of rhyming and word generation with FMRI. *Human Brain Mapping*, *10*(3), 99–106.

McGuire, P., Robertson, D., Thacker, A., David, A. S., Kitson, N., Frackowiak, R. S., & Frith, C. D. (1997). Neural correlates of thinking in sign language. *Neuroreport*, *8*(3), 695–698.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, *115*(2-3), 123–142.

Monzel, M., Mitchell, D., Macpherson, F., Pearson, J., & Zeman, A. (2022). Aphantasia, dysikonesia, anauralia: Call for a single term for the lack of mental imagery–commentary on dance et al. (2021) and hinwar and lambert (2021). *Cortex*, *150*, 149–152. https://doi.org/10.1016/j.cortex.2022.02.002

Morin, A. (2018). *Inner speech: New voices* (P. Langland-Hassan & A. Vicente, Eds.). Oxford University Press.

Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*(3), 376–382.

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14. https://doi.org/10.18637/jss.v060.i02

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, *43*(9), 1–14. https://doi.org/10.18637/jss.v043.i09

Nedergaard, J. S. K., Christensen, M. S., & Wallentin, M. (2021). Valence, form, and content of self-talk predict sport type and level of performance. *Consciousness and Cognition*, *89*, 103102.

Nedergaard, J. S. K., Wallentin, M., & Lupyan, G. (2022). Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychonomic Bulletin & Review*, 1–25.

Owen, W. J., Borowsky, R., & Sarty, G. E. (2004). FMRI of two measures of phonological processing in visual word recognition: Ecological validity matters. *Brain and Language*, *90*(1-3), 40–46.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*(6418), 342–345.

Pedersen, T. L. (2021). *Ggforce: Accelerating 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=ggforce

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Loevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239.

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *13*(5), 687–697.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., et al.others. (1996). Cerebral organization of component processes in reading. *Brain*, *119*(4), 1221–1238.

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roebuck, H., & Lupyan, G. (2020). The internal representations questionnaire: Measuring modes of thinking. *Behavior Research Methods*, *52*(5), 2053–2070.

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.

Soloducha, A. (2020). What it's like living without an inner monologue. CBC News. Retrieved from https://www.cbc.ca/news/canada/saskatchewan/inner-monologue-experience-science-1.5486969

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.

Thiébaut, F. I., White, S. J., Walsh, A., Klargaard, S. K., Wu, H.-C., Rees, G., & Burgess, P. W. (2015). Does faux pas detection in adult autism reflect differences in social cognition or decision-making abilities? *Journal of Autism and Developmental Disorders*, *46*(1), 103–112. https://doi.org/10.1007/s10803-015-2551-1

Wei, T., & Simko, V. (2021). *R package 'corrplot': Visualization of a correlation matrix*. Retrieved from https://github.com/taiyun/corrplot

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2021b). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Bryan, J. (2021). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.* Retrieved from https://CRAN.R-project.org/package=cowplot

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Winsler, A., De Leon, J. R., Wallace, B. A., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behaviour. *Journal of Child Language*, *30*(3), 583–608.

Xie, Y., & Allaire, J. (2022). *Tufte: Tufte's styles for r markdown documents.* Retrieved from https://CRAN.R-project.org/package=tufte

Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, *34*(1), 1–13.

https://doi.org/10.18637/jss.v034.i01

Zeman, A. Z., Della Sala, S., Torrens, L. A., Gountouna, V. E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*, *48*, 145–155.

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra

Table 4

Descriptive statistics of reaction time and accuracy on the task switching experiment.

Group	**Condition**	**Reaction time (ms)**	**95% CI (reaction time)**	**Accuracy**	**95% CI (Accuracy)**
High-verbal	Blocked addition	2287.38	47.04	97.94	0.83
High-verbal	Color-cued switch	2774.63	61.61	95.64	1.16
High-verbal	Blocked subtraction	2527.52	53.77	97.65	0.89
High-verbal	Symbol-cued switch	2564.20	54.44	97.72	0.86
High-verbal	Un-cued switch	2678.94	59.15	94.59	1.29
Low-verbal	Blocked addition	2312.32	46.34	98.32	0.76
Low-verbal	Color-cued switch	2781.48	62.98	95.08	1.26
Low-verbal	Blocked subtraction	2572.91	55.19	97.80	0.88
Low-verbal	Symbol-cued switch	2639.81	56.00	96.72	1.03
Low-verbal	Un-cued switch	2709.74	63.84	93.19	1.47