Not everybody has an inner voice: Behavioral consequences of anendophasia

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

[1] Department of Nordic Studies and Linguistics, University of Copenhagen

[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, Emil Holms Kanal 2, 2300 Copenhagen S, Denmark. E-mail: jskn@hum.ku.dk

Abstract

It is commonly assumed that inner speech – the experience of thought as occurring in a natural language – is a human universal. Recent evidence, however, suggests that the experience of inner speech in adults varies from near constant to non-existent. We propose a name for a lack of the experience of inner speech – anendophasia – and report four studies examining some of its behavioral consequences. We found that adults who report low levels of inner speech (N = 46) have lower performance on a verbal working memory task and have more difficulty performing rhyme judgments compared to adults who report high levels of inner speech (N = 47). Task switching performance - previously linked to endogenous verbal cueing - and categorical effects on perceptual judgments were unrelated to differences in inner speech.

Statement of Relevance

Most adults report experiencing an inner voice, and believe that it plays an important role in their daily lives. However, others report that they do not experience such an inner voice. Although these differences are stable, we do not know whether they have any consequences for how people solve problems and act in the world. In this article, we found that adults with less inner speech differed from adults with more inner speech on some tasks that we thought would involve inner speech, but not others. It is important to understand such individual differences in inner speech use because it has consequences for how we discuss the role of inner speech generally in human life.

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

Not everybody has an inner voice: Behavioral consequences of anendophasia

Research Transparency Statement

Conflicts of interest: Both authors declare no conflicts of interest. Funding: This research was supported by NSF-PAC 2020969 to G.L. Artificial intelligence: No artificial intelligence assisted technologies were used in this research or the creation of this article. Ethics: This research complies with the Declaration of Helsinki (2023), aside from the requirement to preregister human subjects research, and received approval from the Institutional Review Board at the University of Wisconsin-Madison. Open science: The experiment code, materials, data, and analysis scripts can be accessed at https://github.com/johannenedergaard/anendophasia. The studies were not preregistered.

1 Introduction

Everyone, it is often said, has an inner voice: 'Daily, human beings are engaged in a form of inner dialogue, which enables them to [engage in] high-level cognition, including self-control, self-attention and self-regulation.': (Chella & Pipitone, 2020, p. 287); 'We all hear a voice inside our brain, commonly called "inner voice", "inner speech" or referred to as "verbal thoughts" ' (Perrone-Bertolotti, Rapin, Lachaux, Baciu, and Loevenbruck (2014), p. 22). Most people do report experiencing inner speech (Alderson-Day & Fernyhough, 2015; Heavey & Hurlburt, 2008; Morin, Duhnych, & Racy, 2018) and because we often assume that our experiences mirror those of others, the majority experience comes to be viewed as universal (Lupyan, Uchiyama, Thompson, & Casasanto, 2023). The assumption that everyone has an inner voice has served as a stepping stone for research into the functions of inner speech – if everyone has it, it must be important. Speculations have ranged from the idea that natural language constitutes (at least some types of) thought (Bermúdez, 2007; Carruthers, 2002; Clark, 1998; Frankish, 2018; Gauker, 2011) or is necessary for self-awareness (Morin, 2018) to investigations of connections between inner speech and specific processes such as cognitive control (Alderson-Day & Fernyhough, 2015; Cragg & Nation, 2010; e.g., Emerson & Miyake, 2003; Morin et al., 2018), behavioral control (e.g., Nedergaard, Christensen, & Wallentin, 2023), and planning and problem-solving (Lidstone, Meins, & Fernyhough, 2010; e.g., Morin et al., 2018; Wallace, Peng, & Williams, 2017)[1]. But not everyone experiences inner speech. This is attested by personal narratives such as 'What it's like living without an inner voice' (Soloducha, 2020) and 'People With No Internal Monologue Explain What It's Like

[1] We use the terms "inner speech" and "inner voice" interchangeably, but we are not committed to the view that inner speech has all the same auditory and articulatory features as overt speech (Fernyhough & Borghi, 2023, for a recent overarching review; Langland-Hassan, 2018). Importantly, inner speech displays variation both in terms of its form (e.g., dialogic vs. condensed) and modality (e.g., inner speech as hearing a voice vs. experiencing the imagined articulation of speech) (Alderson-Day, Mitrenga, Wilkinson, McCarthy-Jones, & Fernyhough, 2018; Grandchamp et al., 2019; Gregory, 2016; Perrone-Bertolotti et al., 2014). There is evidence that the different modalities of inner speech involve different neural and cognitive mechanisms (e.g., Nalborczyk et al., 2023; Tian, Zarate, & Poeppel, 2016).

In Their Head' (Felton, 2020), as well as more systematic investigations both targeting variation in inner speech (Alderson-Day et al., 2018; Brinthaupt, 2019; Hurlburt, Heavey, & Kelsey, 2013) and auditory imagery, which has sometimes been used as a proxy for inner speech (Dawes, Keogh, Andrillon, & Pearson, 2020; Hinwar & Lambert, 2021).

1.1 The Present Study

We recruited participants differing in subjectively reported inner speech and tested them on four behavioral tasks. These tasks were chosen based on prior theoretical claims that suggested performance on them may differ as a function of inner speech. First, just as visual imagery has been predicted (and sometimes found) to be linked to visual memory, we tested whether inner speech predicted memory for verbal material. We focused on memory for sets of words that were either phonologically similar and orthographically different or orthographically similar and phonologically different. Less inner speech was predicted to be associated with poorer overall memory for verbal material, but to the extent that phonological similarity creates memory confusion (Baddeley, 1966; Murray, 1968), less inner speech may be associated with a reduced phonological similarity effect. Second, participants completed a rhyme judgment task: participants saw pairs of images and needed to indicate whether their names rhymed or not. We reasoned that although participants with low inner speech would have no trouble naming the objects, a reduced reliance on inner speech would make it harder to compare the names in memory – necessary for making a rhyme judgment (Geva, Bennett, Warburton, & Patterson, 2011; Langland-Hassan, Faries, Richardson, & Dietz, 2015). Third, there is substantial evidence that inner speech is often recruited for behavioral control when participants have to switch between different tasks (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Laurent et al., 2016; Miyake, Emerson, Padilla, & Ahn, 2004). For example, when asked to switch between adding and subtracting numbers, participants show a selective impairment if they undergo articulatory suppression, but no such impairment is found if the cues are exogenously provided, e.g., a symbol or color cue is used to inform participants whether they should add or subtract (see Nedergaard, Wallentin, & Lupyan, 2022, for a systematic review of verbal interference effects). We reasoned that people who do not habitually use inner speech might be selectively impaired when they have to rely on self-generated cues to keep track of which task they should be doing. On the other hand, it is possible that they have learned to rely on other strategies in which case no difference would be found. Our fourth task involved examining category effects in perception. There is considerable evidence that language induces more categorical representations from basic perception onward (Forder & Lupyan, 2019; Perry & Lupyan, 2014; e.g., Winawer et al., 2007). In a study examining the effects of conceptual categories, Lupyan, Thompson-Schill, and Swingley (2010) showed that - controlling for visual differences - people's ability to tell whether two stimuli were physically the same was affected by the categorical status of those stimuli. For example, it took longer to distinguish two cats than an equally visually similar cat and dog. We wondered whether such category effects, insofar as they may be in part induced by feedback from verbal labels, may be reduced in people with less inner speech. For all four experiments, we were also interested in whether performance differed by whether participants reported talking out loud during the task.

2 Open Practices Statement

The experiment code, materials, data, and analysis scripts can be accessed at https://github.com/johannenedergaard/anendophasia. The studies were not preregistered.

3 Methods

3.1 Measurement of inner speech

We measured subjectively experienced inner speech using a previously developed and validated Internal Representations Questionnaire (IRQ) (Roebuck & Lupyan, 2020). This questionnaire is broadly similar to other surveys of inner-speech (e.g., the General Inner Speech Questionnaire: Racy, Morin, & Duhnych, 2020; the Self-Talk Scale: Brinthaupt, Hein, & Kramer, 2009; the Varieties of Inner Speech Questionnaire:

McCarthy-Jones & Fernyhough, 2011), and its verbal factor is most closely related to dialogic inner-speech as measured by the Varieties of Inner Speech Questionnaire (r~.7). Two advantages of the IRQ are that its inner speech questions are more inclusive than those on the other scales and the same instrument can be used to assess other individual differences such as visual and orthographic imagery. As is true of other scales, the IRQ measures propensities rather than abilities. Geva and Warburton (2019) suggested that inner speech could be objectively measured using behavioral tasks like silent-rhyme judgments, however the authors did not actually show whether differences in inner speech are associated with differences in performance on silent-rhyme judgment tasks - a limitation we address in the current work.

3.2 Participants

Prior to beginning the study, we had administered the IRQ to university undergraduates and crowdworkers on Amazon Mechanical Turk as part of unrelated studies. From this original pool of 1037 participants, we contacted participants with scores <= 3.5 (bottom 30%-ile) or >= 4.25 (top 20%-ile) on the Verbal factor of the questionnaire which is largely centered on propensity to experience and rely on inner speech. For example, one item with a high loading on the Verbal factor was ‘I think about problems in my mind in the form of a conversation with myself’. One item with a high loading on the Visual factor was ‘I often enjoy the use of mental pictures to reminisce’ (see Supplemental Materials for all verbal factor items). The percentile cut-offs were asymmetric because the distribution in verbal scores on the IRQ is negatively skewed. Recruiting for example the top and bottom quartiles would have resulted in a “low inner speech” group who had moderate amounts of reported inner speech. The final sample included participants from the bottom 20%-ile and the top 29%-ile (see histogram with cutoff values in Supplemental Materials).[2] We received ethical approval from the Institutional Review Board at the University of Wisconsin-Madison. Ten participants

[2] Due to a recruiting error, three participants recruited for the more inner speech group had verbal scores slightly below 4.25 (4.17).

were excluded for responding randomly, missing at least one experiment, or clearly not complying with task instructions. Our final sample included 47 participants with relatively high verbal factor scores on the IRQ and 46 participants with relatively low verbal factor scores. The two groups were balanced in terms of age, gender, education level, dyslexia, and first language. See Table 1. Due to a technical error, demographic data for one participant in the low inner speech group was missing. We were interested in detecting medium-to-large effects. Our sample size allows us to detect effect sizes of approximately .6 at 80% power or .7 at 91% power (two-tailed t-test of mean difference between two independent groups). Power is lower for the reported interactions and so we urge caution in interpreting them.

3.3 Method: Verbal working memory

3.3.1 Materials and procedure. We used word sets from Baddeley (1966) which were designed to vary in phonological and orthographic similarity, while holding constant other psycholinguistic factors. The phonologically-similar set contained the words “bought”, “sort”, “taut”, “caught”, and “wart”. The orthographically similar set contained the words “rough”, “cough”, “through”, “dough”, and “bough”. The control set contained the words “plea”, “friend”, “sleigh”, “row”, and “board”. On a given trial, participants saw five written words in random order from one of the sets. The words were presented sequentially, see Figure 1. After the last word, participants were asked to type the five words they just saw in the order they saw them. Participants began the task by completing 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). Participants then performed 24 trials in total 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.

3.4 Method: Rhyme judgments

3.4.1 Materials and procedure. We constructed a set of rhyme pairs with 20 orthographic pairs (e.g., “sock” and “clock”) and 20 non-orthographic pairs (e.g., “drawer” and “door”). See Supplemental 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 Rossion and Pourtois (2004) because those image sets contained simple images (objects with no background) that had relatively high name agreement. On each trial, participants saw two images of items presented simultaneously and were asked to judge whether the names of the items rhymed or not. Participants completed 60 rhyme judgments in randomized order (20 orthographic rhymes, 20 non-orthographic rhymes, and 20 no-rhyme control trials). There was a 5000 ms response deadline. See Figure 2.

3.5 Method: Task switching

3.5.1 Materials and procedure. On each block, participants were shown 30 randomly selected integers between 13 and 96 and asked to add or subtract 3 from each. All participants completed five blocks beginning with blocked addition or blocked subtraction, followed by (in a counterbalanced order) a block where problems alternated between addition and subtraction with the operation marked by color (red/blue), marked with a symbol (+/-), or not marked. The unmarked block required participants to remember which operation they had just done. 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 3.

3.6 Method: Same/different judgments

3.6.1 Materials and procedure. This experiment used three different black silhouettes of cats and three different black silhouettes of dogs. Participants completed two blocked conditions: making physical identity judgments (same means physically identical) and making category judgments (same means same category). We are only interested in the physical identity judgments here. Participants completed 200 total trials and received feedback after incorrect responses ('incorrect' in red font). See Figure 4.

3.7 Method: Questionnaire

After completing the four experiments, participants answered a series of questions about their experience with inner speech (e.g. 'How often do you have songs stuck in your head?' and '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?') and completed the Varieties of Inner Speech Questionnaire-Revised (VISQ-R) (Alderson-Day et al., 2018). The VISQ-R measures the extent to which inner speech is experienced as dialogic (e.g., 'I talk back and forward to myself in my mind about things') and condensed (e.g. 'My thinking in words is shortened compared to my normal out-loud speech') as well as whether the participant experiences the voices of other people. The questionnaire also measures the perceived functions of inner speech through asking about inner speech as an evaluative and regulatory tool (e.g., ‘I think in inner speech about what I have done, and whether it was right or not’). See Supplemental Materials for the full set of custom questions.

3.8 Data analysis

All analyses were conducted in R version 4.1.3 (R Core Team, 2022). Participants and items (where appropriate) were modeled as random intercepts; random slopes were included for within-subject factors unless it prevented convergence. All predictors were centered. Reaction times were log-transformed to yield a more normal distribution. Accuracies were modeled using logistic regression. For ease of interpretation, the figures show the two inner speech groups as distinct, but all the statistical models use verbal score (average score on the verbal representation items on the Internal Representations Questionnaire) as a continuous predictor. Error bars on all figures represent within-participant 95% confidence intervals around the mean (adjusted for repeated measures). All four experiments were conducted using custom-written software with the JavaScript package jsPsych version 6 (De Leeuw, 2015), and data and code can be found at https://github.com/johannenedergaard/anendophasia.

4 Results

4.1 Verbal working memory

In the verbal working memory experiment, we tested whether the number of words that participants were able to correctly recall (the dependent variable) was predicted by participants’ verbal score on the IRQ and the type of word set (control set, orthographic similarity set, phonological similarity set).

4.1.1 Descriptive statistics by group: Verbal working memory.

Participants with more inner speech 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 5).

4.1.2 Statistical models: Verbal working memory.

Participants remembered phonologically similar words significantly worse (M = 3.22) than orthographically-similar words (M = 3.62) (β = -0.72; SE = 0.08; t = -8.84; p < .001; standardized β = -0.22 [-0.33,-0.11]) which were in turn remembered worse than the dissimilar words (M = 3.94) (β = -0.33; SE = 0.08; t = -3.98; p < .001; standardized β (effect size) = -0.47 [-0.57,-0.36]). Collapsing across the three types of word lists, greater inner speech was associated with better performance (β = 0.27; SE = 0.10; t = 2.60; p = .011; standardized β (effect size) = 0.17 [0.04,0.31]). This effect remained significant when we ignore the recalled order of the words, counting only whether they recalled the correct words (β = 0.19; SE = 0.08; t = 2.57; p = .012; standardized β (effect size) = 0.18 [0.04,0.32]). There were no interaction effects (all p > .10), although numerically, the effect of inner speech was smallest for orthographically similar words (see Figure 5).

4.1.3 Strategies: Verbal working memory.

The groups with more and less inner speech were similar in their reported use of talking out loud as a strategy for remembering the words: 10 out of 47 in the group with more inner speech; 13 out of 46 in the group with less inner speech ($\chi^2(1)$ = 0.29, p = .59). Nevertheless, talking out loud was associated with performance in different ways between the two groups (see Figure 6). As the figure indicates, there was an interaction effect between talking out loud and verbal score on recall (β = -0.50; SE = 0.23; t = -2.19; p = .031; standardized β (effect size) = -0.14 [-0.26,-0.01]). Participants with less inner speech who reported using overt language during the task performed similarly to participants with more inner speech, suggesting that what mattered for performance was the use of speech, either covert or overt.

4.2 Rhyme judgments

In the rhyme judgment experiment, we tested whether the speed and accuracy with which participants made rhyme judgments (the dependent variables) were predicted by participants' verbal score on the IRQ and the type of rhyme (orthographic rhyme, non-orthographic rhyme, and no rhyme). We also tested whether participants' rhyme judgment performance differed by whether they reported talking out loud to remember the words. Five image pairs of rhyming objects – bin/chin, cab/crab, rake/cake, wave/cave, and park/shark – were incorrectly judged to not rhyme on at least half the trials. This was most likely because participants did not name one or both of the images with the intended names (mean agreement rating for these 10 images = 0.58; range = 0.05 to 1). We therefore excluded these trials from further analysis. In addition, we trimmed reaction times below 200 ms (68 trials, 1.4%).

4.2.1 Descriptive statistics by group: Rhyme judgments. Participants who reported having more inner speech were numerically both faster and more accurate than participants who reported having less inner speech on all three types of trials, see Table 3, and Figure 7.

4.2.2 Statistical models: Rhyme judgments. Participants took longer to make rhyme judgments on no-rhyme trials (M = 1981) compared to orthographic trials (M = 1730) (β = 0.12; SE = 0.04; t = 3.01; p = .005; standardized β (effect size) = 0.1 [0.04,0.17]). Non-orthographic trials (M = 1823) did not differ significantly from orthographic trials (β = 0.05; SE = 0.04; t = 1.18; p = .24; standardized β (effect size) = 0.04 [-0.03,0.11]). Higher name agreement was associated with faster RTs (β = -0.04; SE = 0.02; t = -2.22; p = .031; standardized β (effect size) = -0.03 [-0.06,0]). Reported inner speech had no effect on speed of correct rhyme judgments (β = -0.02; SE = 0.02; t = -0.82; p = .42; standardized β (effect size) = -0.01 [-0.05,0.02]). There were no interactions between rhyme type and inner speech (both p's $>$ .31) or between inner speech and the effect of name agreement on accuracy (p $>$ .96).

Participants were more accurate when judging no-rhyme trials as not rhyming (M = 97.03%) than on orthographic rhyme judgments (M = 88.28%) (β = 1.67; SE = 0.32; z = 5.15; p $<$ .001; standardized β (effect size) = 1.67 [1.04,2.31]) and were less accurate on non-orthographic rhyme judgments (M = 80.8%) than on orthographic rhyme judgments (β = -0.59; SE = 0.28; z = -2.07; p = .039; standardized β (effect size) = -0.59 [-1.14,-0.03]). Importantly, a higher verbal score was associated with greater accuracy (β = 0.34; SE = 0.12; z = 2.81; p = .005; standardized β (effect size) = 0.34 [0.1,0.58]). Name agreement did not affect accuracy (p $>$ .13). There were no significant interactions between rhyme type and inner speech (both p > .22) or between inner speech and effect of name agreement on accuracy (p = .51).

4.2.3 Strategies: Rhyme judgments. Asked about their strategies, similar proportions of participants in both groups reported naming the pictures out loud: 23 out of 47 in the higher inner speech group and 21 out of 46 in the lower inner speech group ($\chi^2(1) = 0.01$, p = .91). We observed a similar interaction here as with the memory task (compare Figure 6 and Figure 8). For people who did not report speaking out loud, less habitual inner speech was associated with lower accuracy; for people who did, it was not (orthographic rhymes: β = -0.78; SE = 0.36; z = -2.19; p = .028; standardized β (effect size) = -0.39 [-0.74,-0.04]; non-orthographic rhymes: β = -0.75; SE = 0.34; z = -2.24; p = .025; standardized β (effect size) = -0.37 [-0.7,-0.05]), suggesting once again that speech use - whether covert or overt - is associated with higher accuracy.

4.3 Task switching

In the task switching experiment, we tested whether the speed and accuracy of performing simple arithmetic operations (adding and subtracting) were predicted by participants' reported inner speech (verbal score on the IRQ) as a function of how they were cued to alternate between two operations: addition and subtraction when cued to the correct operation by a symbol, by a color, or having to rely on their memory of which operation they just did. We excluded trials with RTs over 10 seconds (0.5 % of trials). We also recalculated the accuracy measure so that a failure to switch did not render all subsequent trials incorrect as long as the participant proceeded to switch appropriately and obtain the arithmetically correct answer.

4.3.1 Descriptive statistics: Task switching. As can be seen from Table 4 and Figure 9, accuracy was high in all conditions, and reaction times were comparable across the two groups of participants.

4.3.2 Statistical models: Task switching. Participants responded less accurately in the symbol-cued switch condition (M = 97.2%), in the color-cued switch condition (M = 95.4%), and in the un-cued switch condition (M = 93.9%) compared to the blocked addition condition (M = 98.1%) (addition versus symbol-cue: β = -0.42; SE

= 0.18; z = -2.32; p = .020; standardized β (effect size) = -0.42 [-0.77,-0.07]; addition versus color-cue: β = -0.97; SE = 0.17; z = -5.84; p < .001; standardized β (effect size) = -0.97 [-1.3,-0.65]; addition versus un-cued: β = -1.27; SE = 0.16; z = -7.92; p < .001; standardized β (effect size) = -1.27 [-1.59,-0.96]). Accuracy did not differ between blocked subtraction (M = 97.7%) and blocked addition (p = .24). More inner speech was not associated with different accuracy (p = .55) and there were no interaction effects between inner speech and block-type (all p's > .07). Numerically, verbal score interacted with the un-cued condition and cancelled out the very slight (non-significant) reaction time advantage of a higher verbal score.

Participants responded faster in the blocked addition condition (M = 2300 ms) compared to the subtraction condition (M = 2550 ms) (β = 0.09; SE = 0.01; t = 8.41; p < .001; standardized β (effect size) = 0.08 [0.06,0.1]), the symbol-cued switch condition (M = 2601 ms) (β = 0.12; SE = 0.01; t = 9.69; p < .001; standardized β (effect size) = 0.1 [0.08,0.13]), the color-cued switch condition (M = 2778 ms) (β = 0.19; SE = 0.02; t = 12.23; p < .001; standardized β (effect size) = 0.17 [0.14,0.19]), and the un-cued switch condition (M = 2694 ms) (β = 0.15; SE = 0.02; t = 9.39; p < .001; standardized β (effect size) = 0.13 [0.11,0.16]). More reported inner speech did not predict reaction times (p = .81), and there were no interaction effects (all p's > .51).

4.3.3 Strategies: Task switching. There was no significant difference between how many participants with more inner speech (20 out of 47) and how many participants with less inner speech (13 out of 46) reported that they had talked to themselves out loud during the task switching experiment ($\chi^2(1) = 1$, p = .32). There were no obvious differences between the effects that talking out loud had on these two groups (see accuracy and reaction time Figure 10).

4.4 Same/different judgments

In the same/different judgment experiment, we tested whether the speed with which participants made correct same/different judgments was predicted by participants' verbal score on the IRQ and the type of judgment (same category of animal or same image). We excluded trials with RTs above 5 seconds (0.7 %) and below 200 ms (0.07 %). Overall accuracy was high, 95.53, and did not differ between the two inner speech groups (95.58 %) and the group with less inner speech (95.48 %). In subsequent RT analyses, we only include correct trials.

4.4.1 Statistical models: Same/different judgments. 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 participants with more inner speech to be slower to make correct 'DIFFERENT' responses when both stimuli where from the same category but physically different (i.e., dog_1 versus dog_2). Within-category trials were generally associated with significantly slower reaction times (M = 923 ms) than between-category trials (M = 843 ms) (β = -0.08; SE = 0.01; t = -7.71; p < .001; standardized β (effect size) = -0.09 [-0.11,-0.06])). See Figure 11. However, there was no interaction between level of inner speech and category-type: (interaction effect: β = 0.00; SE = 0.01; t = -0.06; p = .95; standardized β (effect size) = 0 [-0.02,0.02]).

4.4.2 Strategies: Same/different judgments. There was no significant difference between how many participants with more inner speech (9 out of 47) and how many participants with less inner speech (4 out of 46) reported that they had talked to themselves out loud during the task ($\chi^2(1) = 1.33$, p = .25). There were no differences between the effects that talking out loud had on these two groups.

4.5 Intertask correlations

In addition to finding (or not finding) differences in task performance as a function of inner speech, it is often informative to see whether correlations between tasks and conditions show a different pattern in people with more vs. less inner speech (Keogh, Wicken, & Pearson, 2021). See Figure 12 for a visualization of how performance on the tasks correlated within the participant groups with more or less inner speech.The dark-blue clusters near the diagonal show that for both groups performance *within* tasks (e.g., RTs on the different types of task-switch trials) was strongly correlated, and similarly so for both groups. When it comes to relationships *between* tasks, however, we found several intriguing differences: Participants with less inner speech showed a positive correlation between verbal recall accuracy and non-orthographic rhyme accuracy (r=.48). This group also showed moderate correlations (r's between .3 and .5) between uncued task-switch accuracy and various measures of verbal recall accuracy. In contrast, participants with more inner speech showed weaker relationships between these measures (r's between .16 and .30).

4.6 Questionnaire measures

Responses to many of the included questions differed substantially as a function of inner speech[3]. For reasons of space, however, we only report a few selected ones here (see Supplemental Materials for further correlations). The questions with the clearest differences concerned rehearsing and revising conversations where the participants with more inner speech reported doing so much more often than the participants with less inner speech did (revise past conversation: $t(87.95) = 5.93$; $p < .001$; practice future conversation: $t(89.33) = 5.33$; $p < .001$). Of the VISQ factors, the IRQ verbal representation score was mostly related to the dialogicality of inner speech ($r(90) = .70$; $p < .001$).

Participants who reported more inner speech estimated that more people experience their thoughts in the form of a conversation with themselves ($\beta = 5.08$; SE = 2; $t = 2.55$; $p = .013$; standardized β (effect size) = 0.26 [0.06,0.46]) and that more people hear words in their "mind's ear" when they read ($\beta = 5.09$; SE = 2.07; $t = 2.46$; $p = .016$; standardized β (effect size) = 0.25 [0.05,0.45]). They did not, however, estimate that more people were able to see vivid images in their "mind's eye" ($\beta = 1.17$; SE = 2.25; $t = 0.52$; $p = .61$; standardized β (effect size) = 0.05 [-0.15,0.26]).

5 Discussion

Our study is, to our knowledge, the first to conduct a systematic investigation of whether differences in inner speech have behavioral consequences. Participants who report experiencing less inner speech (our sample targeted those at < 20%ile of the verbal score on the Internal Representations Questionnaire) performed worse when judging whether the names of two images rhymed, and they had poorer verbal working memory. Interestingly, in both the rhyming experiment and the verbal working memory experiment, performance differences between the two groups disappeared when

[3] Data from one participant was missing so we report questionnaire data from 47 participants with more inner speech and 45 participants with less inner speech.

participants reported talking out loud to solve the problems, suggesting that the efficacy of using covert and overt speech in these cases were equivalent. Inner speech differences did not predict performance in task switching, indicating that while inner speech can be used as a behavioral self-cue, other and equally effective strategies may be available. Lastly, categorical effects on perceptual discrimination were similar for the two groups, suggesting either that the categorical effects in such tasks are not language-based, or that the speeded nature of such tasks lessens reliance on inner speech (and language more generally). Examination of intertask correlations indicated that participants with less inner speech were more likely to rely on a common mechanism when performing rhyme judgments, verbal recall, and task switching without an extrinsic cue (requiring them to remember what they just did). Our finding of stronger intertask correlations for participants with less inner speech is conceptually similar to Keogh et al. (2021)'s finding of stronger relationships between different visual working memory tasks in participants with aphantasia compared to those with typical visual imagery.

5.1 Anendophasia: A Lack of Inner Speech

People's self-reports cannot always be taken at face value (Heavey & Hurlburt, 2008; Hurlburt, 2011; Hurlburt et al., 2013). But when people report that they rarely or never experience inner spech, they are not just confabulating. This is evident both in the consistency of their subjective responses (Roebuck & Lupyan, 2020), and, as we report here, differences in objective performance. When investigating unusual human experiences, it helps to have a label. For example, the coining of "aphantasia" to the lack of visual imagery (Zeman et al., 2010) is both helpful for research – providing a useful keyword – and for self-identification; its introduction led to the creation of an online community with over 50,000 members (r/aphantasia). We would therefore like to propose a name for the phenomenon of a lack of inner speech: **anendophasia**: *an* (lack) + *endo* (inner) + *phasia* (speech). This term was developed in consultation with individuals who identify as lacking inner speech and has the benefit of including the familiar Greek root *phasia* (aphasia, paraphasia, etc.). Furthermore, "endophasia" has precedent in being used to refer to inner speech (Bergounioux, 2001; Loevenbruck et al., 2018). The term also avoids subsuming inner speech under “aphantasia” (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 debated) and because the linguistic properties of inner speech are likely not reducible to auditory and articulatory features. For these reasons, we also do not believe the previously proposed term “anauralia” is appropriate (Hinwar & Lambert, 2021).

5.2 Relations to Visual Imagery, Auditory Imagery and “Unsymbolized” Thought

Can anendophasia be thought of simply as a lack of auditory imagery? We think not. First, many who lack inner speech report being able to engage in musical imagery (although they report “earworms” - intrusive musical imagery - less often than people with typical levels of inner speech). Second, although inner speech is often experienced as having phonological features – one of the reasons people often perceive it as speech (Langland-Hassan, 2018) – it can also involve an articulatory-motor dimension (Geva, 2018; Perrone-Bertolotti et al., 2014). The current work was not designed to investigate the separate contributions of auditory and articulatory dimensions. Paradoxically, some people also claim to experience “wordless” inner speech akin to a series of tip of the tongue states (Hurlburt et al., 2013).

When asked to reflect on what form their thoughts take, people who score low on both inner speech and visual imagery claim that they “think in concepts”. What it means to “think in concepts” without relying on language is not clear. Beyond informal self-reports, the existence of such non-verbal and non-perceptual phenomenal experiences is supported by Descriptive Experience Sampling (DES) (Heavey & Hurlburt, 2008; Hurlburt & Akhter, 2006). When participants are probed at random times and asked to report on their mental states, ~22% of the time their reports are consistent with what Hurlburt and colleagues have called “unsymbolized thinking”. In such episodes, people feel that they think ‘a particular, definite thought without awareness of that thought being conveyed as words, images, or any other symbols' (Heavey & Hurlburt, 2008, p. 802). Unsymbolized thinking is a slippery construct that tends to be defined in terms of what it is not. For example, Hurlburt and Akhter (2008) describe it as 'a thinking, not a feeling, not an intention, not an intimation, not a kinesthetic event, not a bodily event' (p. 1366). A telling example is a participant wondering if her friend will arrive in a car or pickup truck, but not experiencing any words or images; rather, the question is experienced as a single undifferentiated whole.

It is possible that such "unsymbolized thinking" is subserved by the same processes as inner speech, but simply lacks conscious auditory or articulatory features of inner speech (Vicente & Martinez-Manrique, 2016). Alternatively, it may correspond to a genuinely different form of experience in which people entertain more abstract conceptual representations which are less accessible to people with higher levels of inner speech and imagery. Anendophasia as we define it pertains to the subjective experience of inner speech. Our current measurements cannot distinguish whether differences in the experience of inner speech are due to a difference in the cognitive processes subserving inner speech, or a difference in the metacognitive awareness of generated phonetic or articulatory features.

5.3 Limitations

One limitation of our work is its reliance on wholly subjective questions for measuring inner speech. Considering that our focus is on differences in phenomenology, this is appropriate. At the same time, there is reason to be skeptical of people's assessments of their inner experiences. People can be wrong about what they think they experience (Hurlburt & Schwitzgebel, 2011). It would be therefore helpful to supplement subjective assessments with objective ones of the sort becoming possible for differences in visual imagery (Kay, Keogh, Andrillon, & Pearson, 2022). Relatedly, since inner speech is known to vary not just between people, but also across situations (Fernyhough, 2004; Grandchamp et al., 2019; Oppenheim & Dell, 2010), it is worth examining whether people who report having little inner speech experience more of it if placed in situations that, e.g., benefit from verbal rehearsal. We believe our results generalize across age, gender, and educational status, but it is an open question whether any of the relationships we report are specific to English speakers or Westerners.

Another limitation is the remaining possibility that differences we ascribe to inner speech come from third factors such, e.g., a general difference in introspection and/or conscientiousness. Although we cannot rule out all such possible confounds, it is worth noting that differences in inner speech, while correlated with e.g., visual imagery, are dissociable from it. There is also no evidence that inner speech was associated with across-the-board differences in performance. Our effects were specific to certain task and condition combinations. That said, to obtain greater confidence in the specific causal role of inner speech on performance, it may be possible to manipulate it via interference or by instructing participants to use it in specific ways.

A further limitation is that our measurement of inner speech does not distinguish between its prevalence in one's conscious experience and one's ability to control its deployment. In ongoing work, we have found that people who report experiencing more inner speech (measured as in the present studies) report having a harder time shutting it off (anecdotally, one reported benefit of mindfulness meditation is precisely this ability; for someone with less inner speech to begin with, there is less to shut off). At the same time, we suspect that there is a wide range in ability to regulate one's inner speech, perhaps related to more domain-general differences in cognitive control.

Lastly, while the term "anendophasia" connotes *lack* of inner speech, many of the participants in our "low inner speech" group reported having *some* inner speech. Screening a larger group to identify people who do not endorse having *any* inner speech would help in knowing whether the cognitive consequences of having less inner speech are continuous with having none.

6 Conclusion

Some people report not experiencing an inner voice, and these reports appear to be related to measurable differences in behavior. We proposed a name for a lack of inner speech: anendophasia. We found that people who experience less inner speech were worse at making rhyme judgments in response to images and remembering lists of words. Task switching performance was not, however, either slower or less accurate, and there were no differences in category effects on perceptual discrimination. Taken together, our experiments suggest that there are real behavioral consequences of experiencing less or more inner speech, and that these differences may often be masked because people with anendophasia use alternate strategies to achieve similar overall performance.

7 References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965.

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.

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.

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.

Brinthaupt, T. M. (2019). Individual differences in self-talk frequency: Social isolation and cognitive disruption. *Frontiers in Psychology*, *10*, 1088.

Brinthaupt, T. M., Hein, M. B., & Kramer, T. E. (2009). The self-talk scale: Development, factor analysis, and validation. *Journal of Personality Assessment*, *91*(1), 82–92.

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

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.

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.

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.

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

Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, *22*(1), 49–68. https://doi.org/10.1016/j.newideapsych.2004.09.001

Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. *Trends in Cognitive Sciences*.

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123.

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

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.

Geva, S., & Warburton, E. A. (2019). A test battery for inner speech functions. *Archives of Clinical Neuropsychology*, *34*(1), 97–113.

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., … Lœvenbruck, H. (2019). The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, *10*, 2019. https://doi.org/10.3389/fpsyg.2019.02019

Gregory, D. (2016). Inner speech, imagined speech, and auditory verbal hallucinations. *Review of Philosophy and Psychology*, *7*, 653–673.

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

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

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

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

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

Hurlburt, R. 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

Hurlburt, R. T., & Schwitzgebel, E. (2011). *Describing inner experience?: Proponent meets skeptic*. Mit Press.

Kay, L., Keogh, R., Andrillon, T., & Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *Elife*, *11*, e72484.

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

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.

Laurent, L., Millot, J.-L., Andrieu, P., Camos, V., Floccia, C., & Mathy, F. (2016). Inner speech sustains predictable task switching: Direct evidence in adults. *Journal of Cognitive Psychology*, *28*(5), 585–592.

Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology*, *107*(4), 438–451. https://doi.org/https://doi.org/10.1016/j.jecp.2010.06.002

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.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, *21*(5), 682–691.

Lupyan, G., Uchiyama, R., Thompson, B., & Casasanto, D. (2023). Hidden differences in phenomenal experience. *Cognitive Science*, *47*(1), e13239.

McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition*, *20*(4), 1586–1593.

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.

Murray, D. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*(4p1), 679–684.

Nalborczyk, L., Longcamp, M., Bonnard, M., Serveau, V., Spieser, L., & Alario, F.-X. (2023). Distinct neural mechanisms support inner speaking and inner hearing. *Cortex*, *169*, 161–173.

Nedergaard, J. S. K., Christensen, M. S., & Wallentin, M. (2023). Mind over body: Interfering with the inner voice is detrimental to endurance performance. *Psychology of Sport and Exercise*, 102472.

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.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160.

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.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72.

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/

Racy, F., Morin, A., & Duhnych, C. (2020). Using a thought listing procedure to construct the general inner speech questionnaire: An ecological approach. *Journal of Constructivist Psychology*, *33*(4), 385–405.

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

Rossion, B., & Pourtois, G. (2004). Revisiting snodgrass and vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236.

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

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12.

Vicente, A., & Martinez-Manrique, F. (2016). The nature of unsymbolized thinking. *Philosophical Explorations*, *19*(2), 173–187.

Wallace, G. L., Peng, C. S., & Williams, D. (2017). Interfering with inner speech selectively disrupts problem solving and is linked with real-world executive functioning. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3456–3460.

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.

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.

Table 1

Comparisons of demographic characteristics of the group with more inner speech and the group with less inner speech.

Measure	More inner speech	Less inner speech	Test for difference
Age	Mean = 36.91; Median = 37; range = 18-67	Mean = 37.56; Median = 39; range = 18-70	t(88.43) = -0.19; p = .85
Gender	22 female, 25 male	19 female, 26 male	$\chi^2(1) = 0.05$; p = .82
Native English-speaker	47 native speakers, 0 non-native speakers	41 native speakers, 4 non-native speakers	$\chi^2(1) = 2.49$; p = .11
Dyslexia	46 non-dyslexic, 1 self-diagnosed	44 non-dyslexic, 1 self-diagnosed	$\chi^2(1) < 0.01$; p > .99
Education level	12 high school diploma, 14 some college - no degree, 6 associate's degree, 14 bachelor's degree, 1 master's degree	1 less than high school, 14 high school diploma, 8 some college - no degree, 7 associate's degree, 11 bachelor's degree, 2 master's degree, 2 PhD, law, or medical degree	t(84.46) = -0.23; p = .82

Table 2

Descriptive statistics by group in the verbal working memory experiment.

Group	Word set	Score (item and position)	95% CI (item and position)	Score (item only)	95% CI (item only)
More inner speech	Control set	4.19	±0.13	4.51	±0.08
More inner speech	Orthographic similarity set	3.72	±0.14	4.18	±0.1
More inner speech	Phonological similarity set	3.43	±0.16	4.11	±0.1
Less inner speech	Control set	3.69	±0.15	4.17	±0.11
Less inner speech	Orthographic similarity set	3.52	±0.15	4.10	±0.11
Less inner speech	Phonological similarity set	3.02	±0.15	3.81	±0.11

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)
More inner speech	Non-orthographic rhyme	1853	±51	83.75	±2.81
More inner speech	No rhyme	1931	±53	98.45	±1.13
More inner speech	Orthographic rhyme	1719	±55	91.99	±2.37
Less inner speech	Non-orthographic rhyme	1976	±54	77.75	±3.19
Less inner speech	No rhyme	2027	±60	95.57	±1.65
Less inner speech	Orthographic rhyme	1859	±60	84.48	±3.16

Table 4

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

Group	**Condition**	**Mean reaction time (ms)**	**95% CI (reaction time)**	**Accuracy**	**95% CI (accuracy)**
More inner speech	Blocked addition	2287	±47	97.94	±0.83
More inner speech	Color-cued switch	2775	±62	95.64	±1.16
More inner speech	Blocked subtraction	2528	±54	97.65	±0.89
More inner speech	Symbol-cued switch	2564	±54	97.72	±0.86
More inner speech	Un-cued switch	2679	±59	94.59	±1.29
Less inner speech	Blocked addition	2312	±46	98.32	±0.76
Less inner speech	Color-cued switch	2781	±63	95.08	±1.26
Less inner speech	Blocked subtraction	2573	±55	97.80	±0.88
Less inner speech	Symbol-cued switch	2640	±56	96.72	±1.03
Less inner speech	Un-cued switch	2710	±64	93.19	±1.47

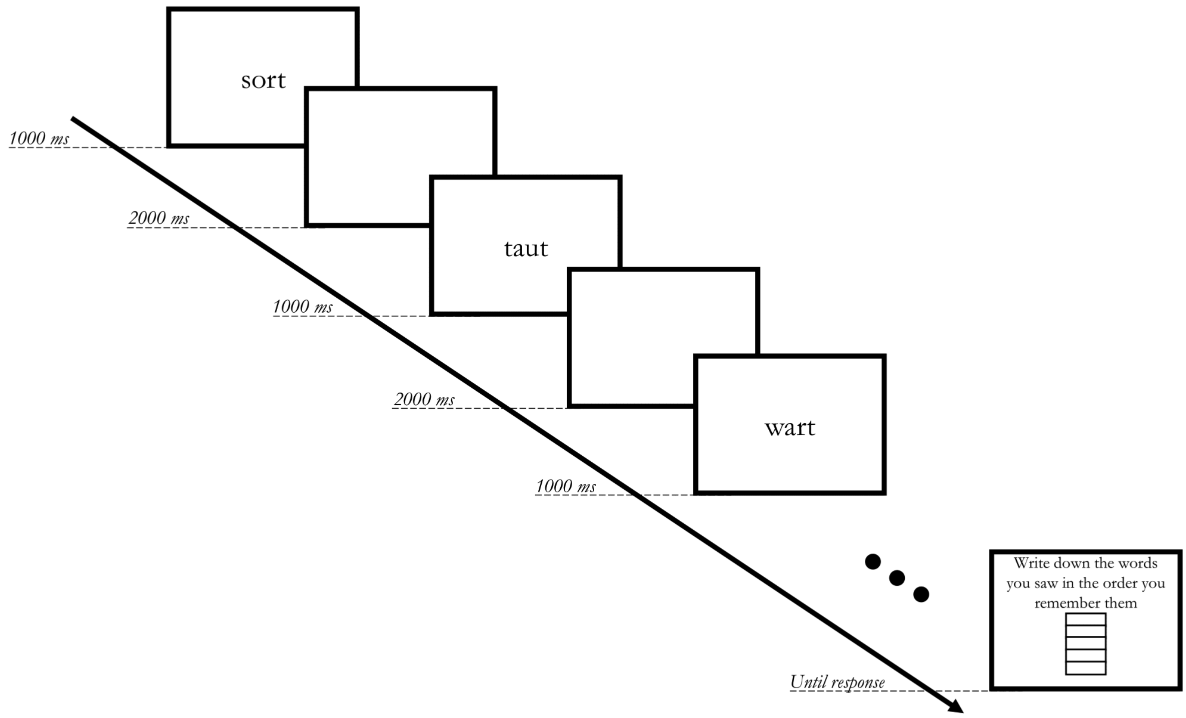

Figure 1. A schematic of the procedure, showing a trial with phonologically-related words. Each trial had five words.

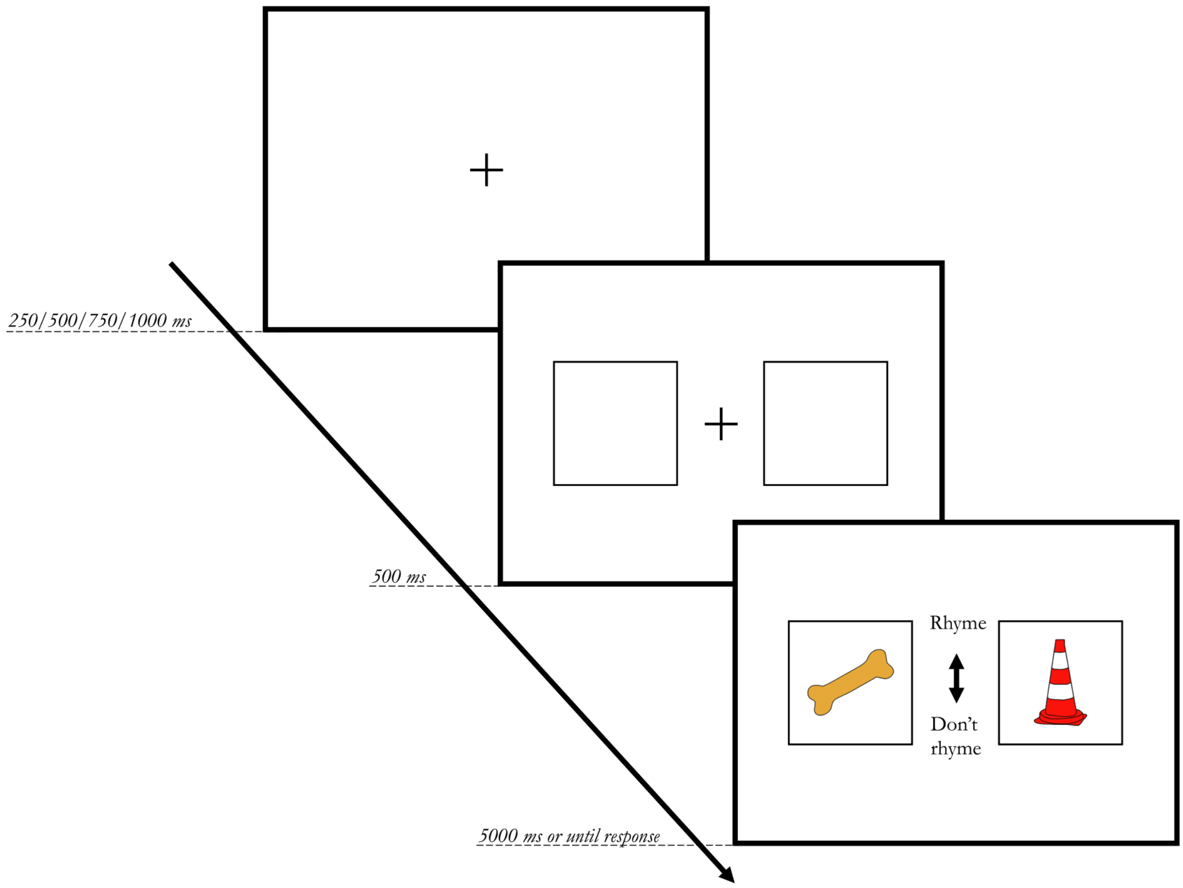

Figure 2. A sketch of a rhyme judgment trial. The stimuli here exemplify an orthographic rhyme – "bone" and "cone" – and the correct answer would therefore be "Rhyme".

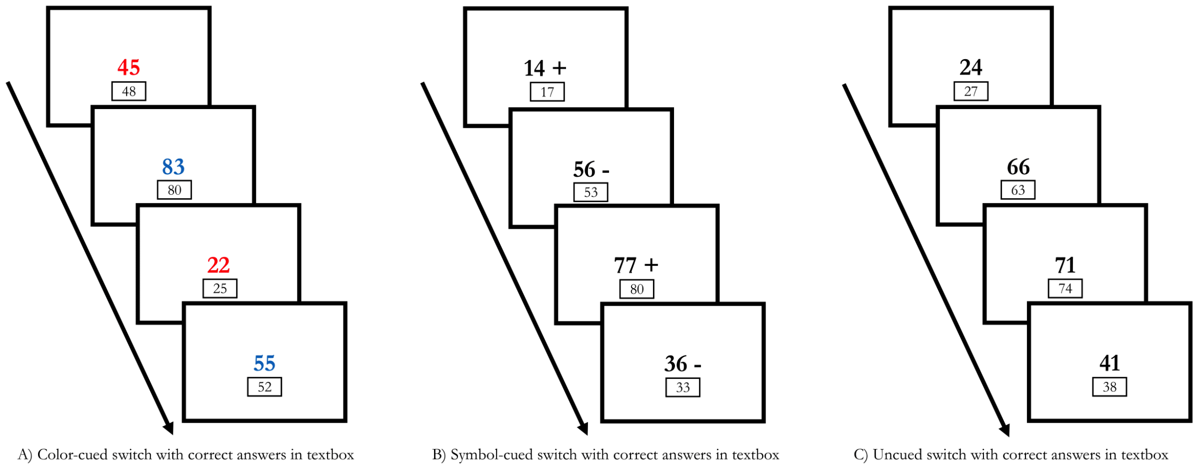

Figure 3. A sketch of the three switched conditions in the task switching experiment. Figure A shows four color-cued switch trials with correct answers, Figure B shows four symbol-cued switch trials with correct answers, and Figure C shows four un-cued switch trials with correct answers.

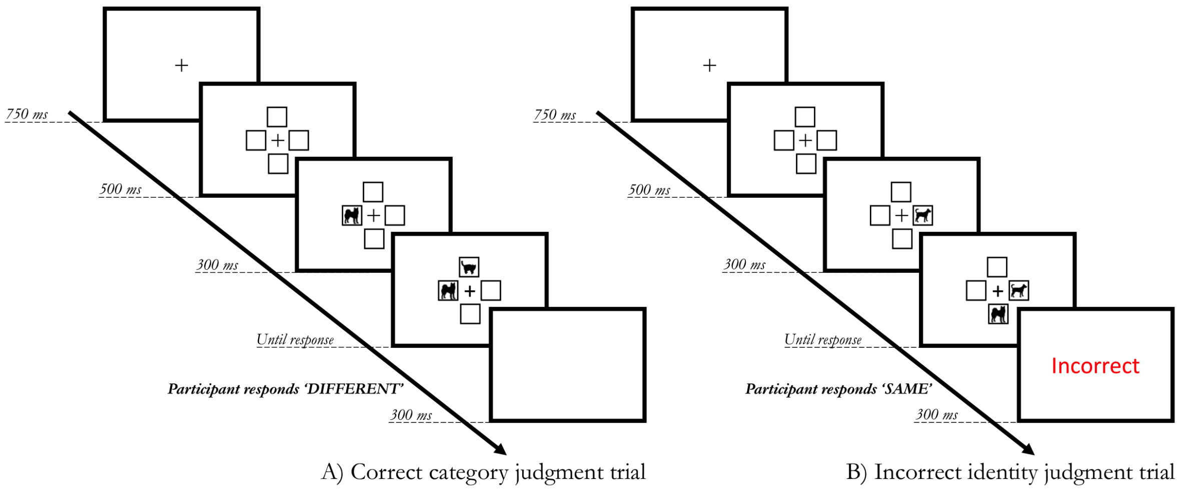

Figure 4. A sketch of the two conditions of the category judgment experiment. On Figure A, we see a correct category judgment trial where the participant responds that the cat and dog silhouettes represent different animals. On Figure B, we see an incorrect identity judgment trial where the participant responds that the two dogs are identical.

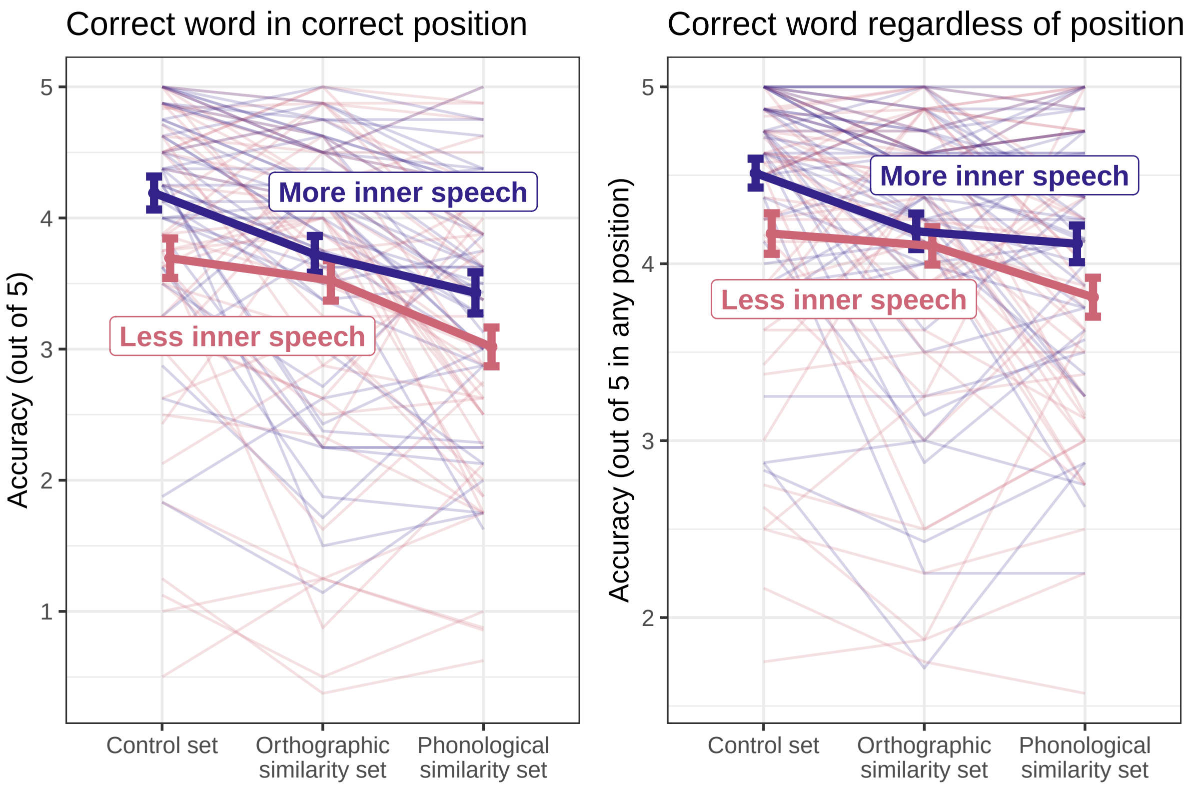

Figure 5. Score on the verbal working memory task by word set.

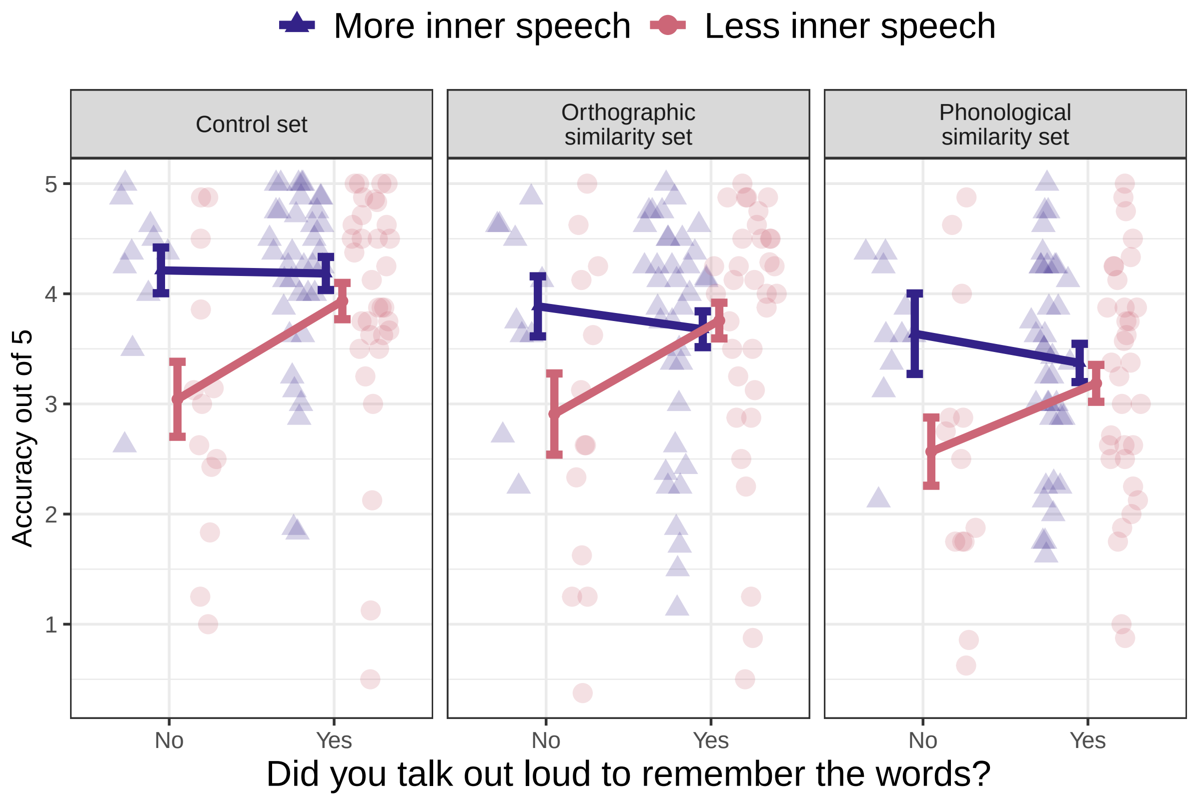

Figure 6. Verbal working memory performance by whether participants reported talking out loud to help them remember or not.

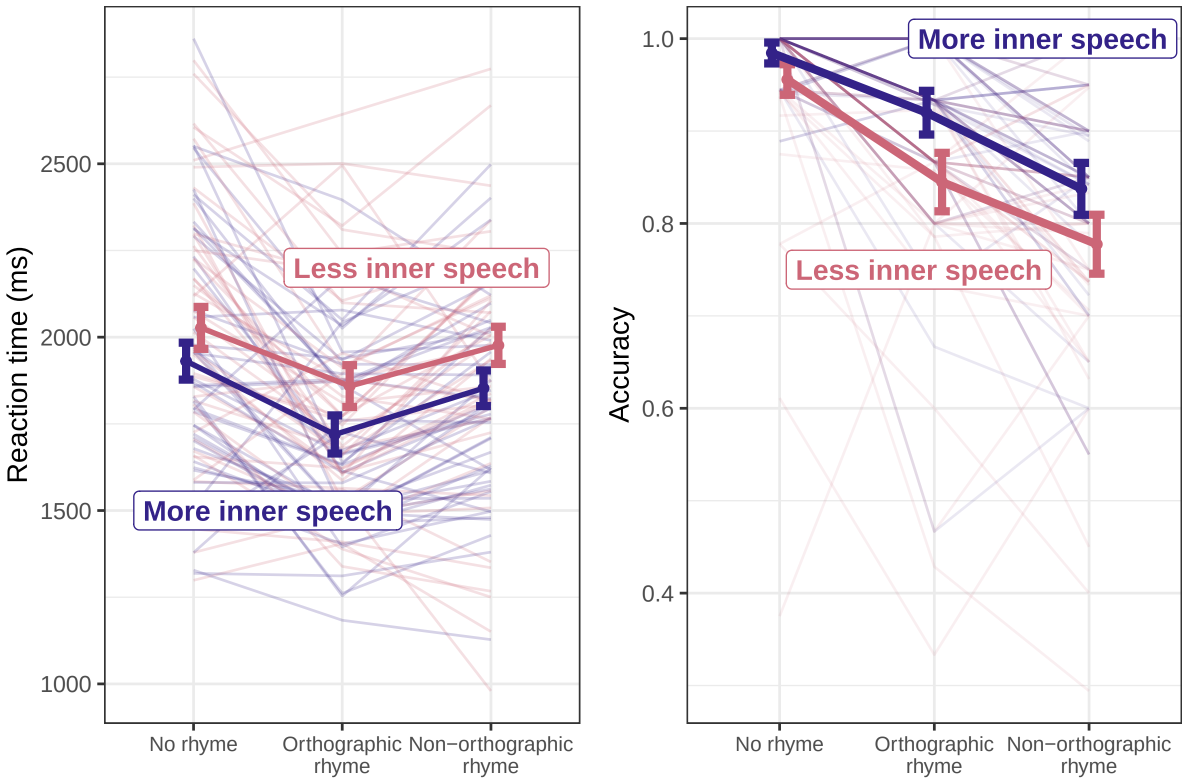

Figure 7. Reaction time and accuracy across groups by rhyme type.

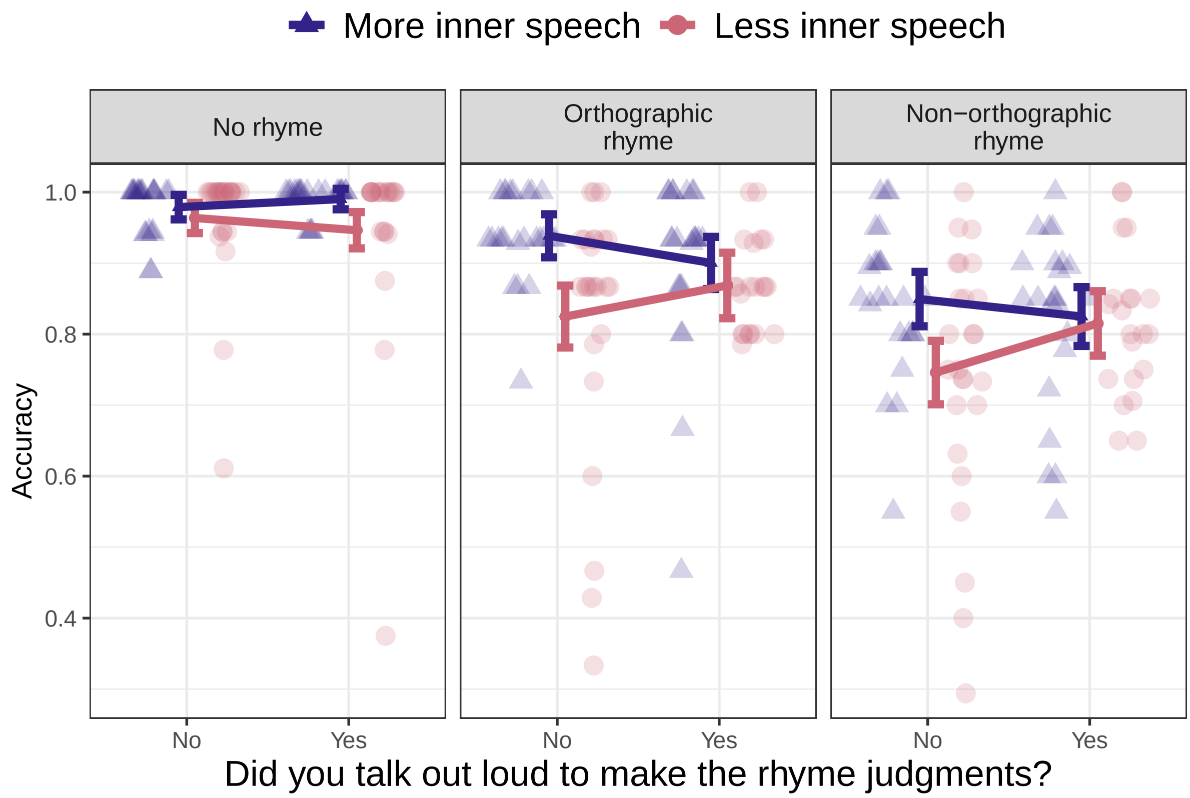

Figure 8. Reaction time and accuracy by whether participants indicated that they had talked out loud to make the rhyme judgments.

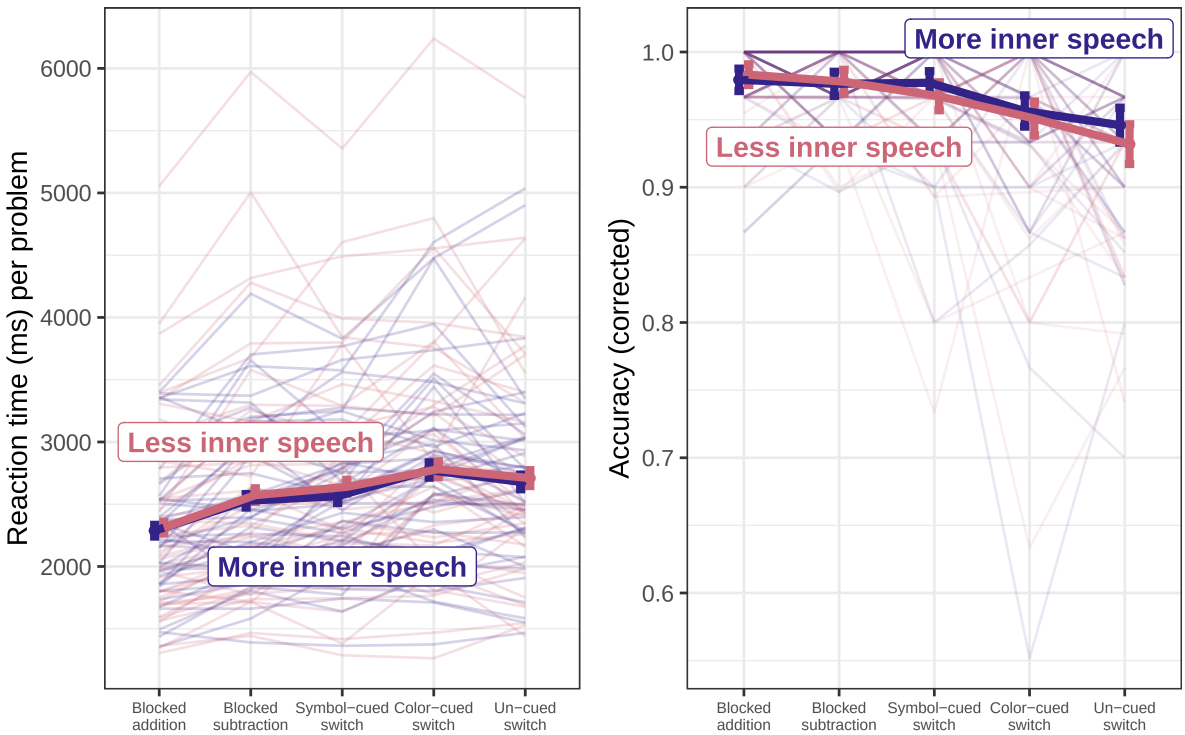

Figure 9. Reaction time and accuracy across conditions in the task switching experiment.

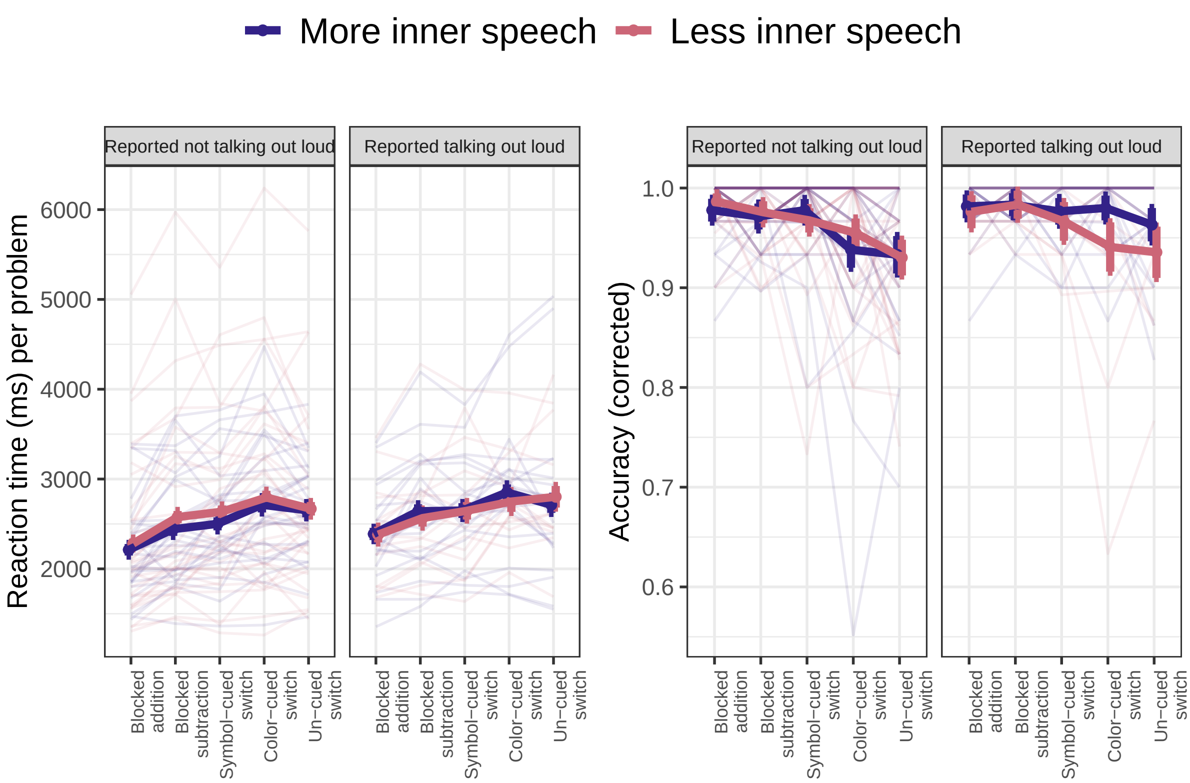

Figure 10. Reaction time (ms) and accuracy in the task switching experiment by whether participants reported talking out loud to remember the correct rule or not.

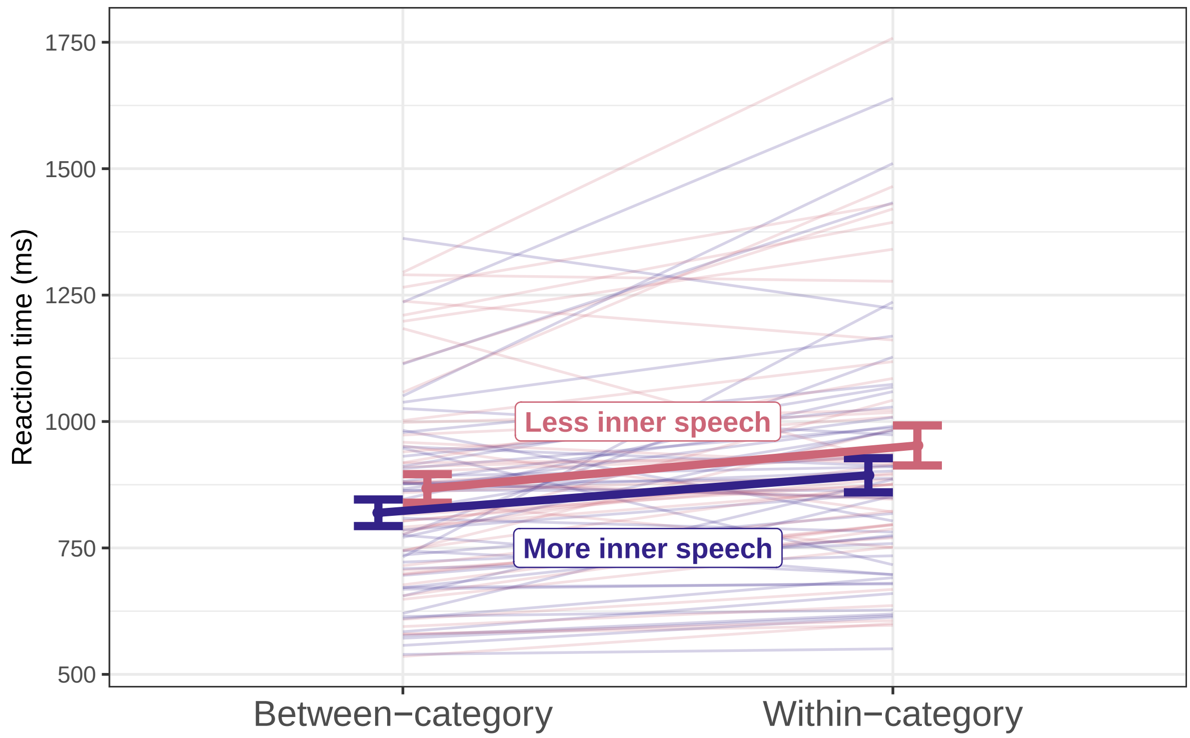

Figure 11. Reaction time on identity trials where the correct response was 'DIFFERENT' either because the two silhouettes were from different categories or different images from the same category.

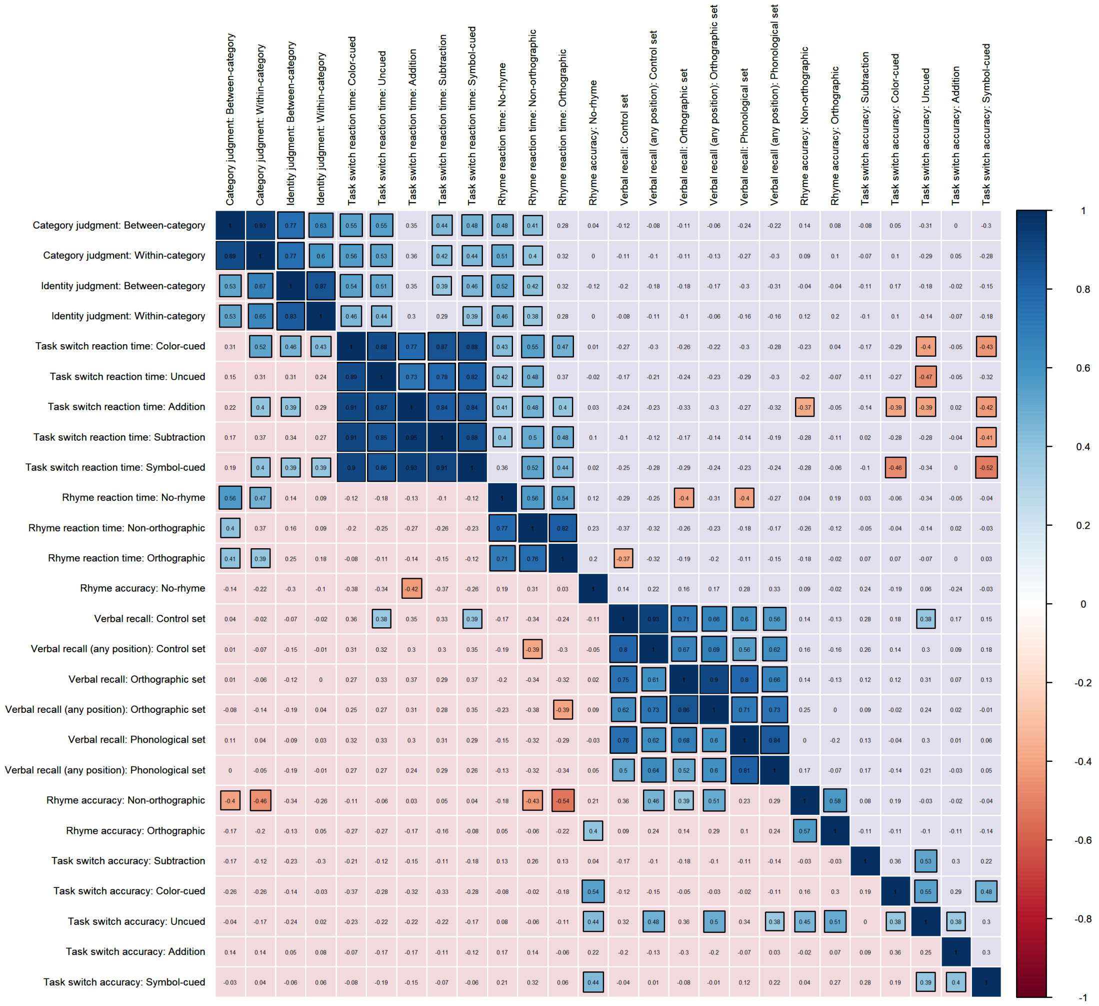

Figure 12. Intertask correlations. The upper triangle shows correlations for the group with more inner speech; the lower triangle shows correlations for the group with less inner speech. Colored squares represent significant correlations at p < .01 (|r|>.38 given the sample size).