Index

Manuscripts

Article I. 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. https://doi.org/10.3758/s13423-022-02144-7

Article II. Nedergaard, J. S. K., Christensen, M. S., & Wallentin, M. (in review). Mind over body: Interfering with the inner voice is detrimental to endurance performance. *PsyArXiv*. Preprint: https://doi.org/10.31234/osf.io/bfj9w

Article III. Nedergaard, J. S. K., Skewes, J. C., & Wallentin, M. (2023). 'Stay focused!': The role of inner speech in maintaining attention during a boring task. *Journal of Experimental Psychology: Human Perception and Performance*. Preprint: https://doi.org/10.13140/RG.2.2.26573.72167

Article IV. Nedergaard, J. S. K. & Lupyan, G. (in review). Not everyone has an inner voice: Behavioral consequences of anendophasia. Preprint: https://github.com/johannenedergaard/anendophasia/blob/main/manuscript/anendophasia_paper.pdf

English summary

Inner speech plays a prominent role in most people's lives where it is used for a variety of things such as encouraging and scolding oneself, planning, and practice for social interactions. Given the perceived prevalence and importance of inner speech, it is important to find out when it helps us and when it does not. The present thesis revolves around questions of the nature and functions of inner speech which I address using a variety of methods.

The thesis begins with an exposition of common questions in inner speech research. First, there is the fundamental question of how we can know that people use inner speech at all. The most straightforward answer is: 'because they report doing so'. There is additional evidence from people's ability to report which natural language their inner speech takes place in and from neuroimaging studies showing that the neural substrates of reported inner speech are similar to speech that is spoken out loud. In addition, we can also sometimes measure micromovements of the articulatory muscles (throat, tongue, etc.) during covert speech. Second, there is the question of how inner and outer speech are related. It seems plausible that inner and outer speech develop in tandem, and that inner speech is generally experienced as an attenuated form of overt speech (shorter, weaker, and less effective for example for supporting memory). The functions of inner and outer speech are likely to be continuous. Third, we need to consider whether inner speech is best characterised as inner *speaking* or inner *hearing*. These two can be differentiated in Descriptive Experience Sampling research where participants carry a beeper and are trained to describe their experience at random times without preconceived notions of what their experience should contain. Inner speaking and inner hearing also appear to have different neural substrates. Inner speech cannot be reduced to either inner hearing (auditory imagery) or inner speaking, although the latter is more plausible since speaking is always accompanied by hearing. The reverse is not the case. Fourth, there is much debate about whether inner speech has all the same

articulatory and phonological features as outer speech, or whether inner speech is abstracted away from the auditory-motor features of speech (sometimes to the extreme that inner speech has no sounds and is 'thinking in pure meanings'). Some theorists see inner speech as being essentially motor simulation and therefore fully specified in terms of phonology and articulatory movements. Others see inner speech as being abstracted away from such specifications but retaining other linguistic properties such as hierarchical structure, semantics, pace, and rhythm. Research so far seems to indicate that the degree of abstractness of inner speech depends on both individuals and situations.

I also outline two theoretical perspectives on the functions of inner speech which served as the starting points for my PhD work: the phonological loop perspective and the Vygotskian self-regulation perspective. According to the phonological loop perspective, inner speech is identical to the phonological loop or verbal working memory, a hypothesised component of working memory where phonological information is refreshed until it is no longer immediately needed (hence "working" memory). There is substantial evidence for a connection between the phonological loop and inner speech although inner speech is likely to serve a wider array of functions than the phonological loop. According to the Vygotskian self-regulation perspective, inner speech has its roots in caregivers' regulatory, child-directed speech which is gradually internalised through development and is used in adults primarily for self-regulation. There is also considerable evidence for this perspective, for example since most children talk out loud to themselves and this appears to help them solve problems.

As part of my MA (integrated with my PhD), I conducted a study of self-talk in athletes. In the first part of that study, we used machine learning methods to predict sport type (badminton or running) from self-talk questionnaire data. In the second part of the study, we again used machine learning methods to predict personal best marathon times from self-talk questionnaire data. The MA thesis itself also focused on combining the Vygotskian self-regulation perspective with the literature on self-talk in sports, and parts of the present PhD thesis are built on that prior work. My focus on the role of inner speech in physical and mental endurance is also partially inspired by the sport psychology literature.

In my PhD work, I combine the Vygotskian and the phonological loop perspectives, using methods from phonological loop research to test Vygotskian hypotheses. Specifically, I rely on the dual-task interference method where we attempt to prevent people from using their inner speech to solve a given task by asking them to perform a secondary, simultaneous verbal task (verbal interference). The logic behind this method is straightforward: If inner speech normally helps people solve a given task, then they should be worse at solving that task when their verbal resources are occupied (compare trying to remember a phone number while simultaneously saying the names of the months out loud). Aside from dual-task interference, I focus on experience sampling and individual differences as methods for investigating inner speech. Using experience sampling, we prompt people at random times to report whether they are experiencing inner speech and what characteristics their inner speech has (content, positive or negative, about the past or the future, etc.) This gives more fine-grained and reliable reports than standard retrospective, general surveys. The individual differences approach lets us explore whether reported differences in inner speech experience are associated with measurable differences in behaviour. This may also shed light on what inner speech is and is not used for.

In **Article I**, I present a systematic review of verbal interference studies. We reviewed 101 studies and found 11 primary task categories. These were categorization (simple and complex), memory, mental arithmetic, motor control, reasoning (verbal materials and non-verbal materials), task switching, theory of mind, visual change, and visuospatial integration and wayfinding. We also found four different kinds of interference tasks (articulatory suppression, memory, verbal shadowing, and judgment tasks). We found evidence for inner speech being involved in tasks where participants needed to remember changing task rules, keep numbers in mind for doing mental arithmetic, and make categorical judgments. However, we did not find convincing evidence for inner speech involvement in reasoning with non-verbal materials, theory of mind, and visuospatial integration and wayfinding. It is important to note that we only examined one way of testing the role of language in a given process, so absence of evidence for interference does not necessarily mean that language is not involved. In the article, we discuss the advantages and disadvantages of verbal interference as an approach to studying the role of inner speech in cognition.

Article II contains two behavioural experiments where participants performed sprints on an exercise bike while under the kind of verbal interference described in Article I. To control for the effect of doing two things at the same time, we compared verbal interference with visuospatial interference. We found that participants were worse at pushing themselves to cycle faster when distracted from talking to themselves, also compared with generally being distracted. We took this to indicate that people normally benefit from talking to themselves to prevent themselves from stopping during endurance tasks. How much participants were affected by interference did not appear to depend on either how much they reported usually talking to themselves or whether they reported that talking to themselves usually helps their performance.

In **Article III**, we combined mind-wandering research (which has tested the detrimental effects of attentional drifts away from an assigned task) and experience sampling to test whether people benefit from talking to themselves when a task is very tedious and demands nothing but their attention. Participants were looking at a blank screen for several minutes at a time and had to respond to intermittent prompts very quickly. After each prompt, participants answered questions about their experience at the time of the prompt (what format it took, whether it was about the task, etc.) We found that participants reacted faster and with less varied responses to infrequently occurring prompts when they were talking to themselves about the task. Just as in Article II, we interpret this as meaning that participants are able to use inner speech to prevent themselves from mind-wandering, i.e., quitting on an attentional endurance task.

In **Article IV**, I present a study of people who experience little to no inner speech and what behavioural consequences this might have. Such participants (with "anendophasia", as we call it) performed less well on tasks specifically requiring phonological working memory (rhyme judgment and immediate serial recall of words) but equally well as the comparison group with more inner speech on visual judgments and task switching between simple addition and subtraction problems. This indicates that self-reported propensity to experience and use inner speech is connected to especially naming, covert sound comparison, and verbal working memory. It remains an open question whether people with anendophasia really experience no inner speech at all or could be said to experience elements of inner speech, e.g. conceptual structure without word labels (i.e., highly abstracted away from auditory and articulatory processes). This is for future studies to explore.

The work presented in the four articles has important implications for how we understand inner speech. The articles underline the intimate connection between inner speech and verbal working memory, help disentangle the specific mechanisms behind the self-regulatory functions of inner speech, and shed new light on individual differences in inner speech experience and its relationship with behaviour.

Dansk resumé

Den indre stemme spiller en fremtrædende rolle i de fleste menneskers liv, hvor den bruges til en række ting såsom at opmuntre sig selv, skælde sig selv ud, planlægge og øve sociale interaktioner. I betragtning af hvor udbredt og afgørende den indre stemme lader til at være, er det vigtigt at finde ud af, hvornår den hjælper os, og hvornår den ikke gør. Denne afhandling kredser om spørgsmål om den indre tales natur og funktioner, som jeg behandler ved hjælp af en række forskellige metoder.

Afhandlingen indledes med en redegørelse for almindelige spørgsmål inden for forskning i den indre stemme. For det første er der det grundlæggende spørgsmål om, hvordan vi kan vide, at mennesker overhovedet bruger indre tale. Det mest ligetil svar er: 'Fordi de rapporterer, at de gør det'. Der er yderligere beviser fra folks evne til at rapportere, hvilket specifikt sprog deres indre tale foregår i, og fra hjerneskanningsundersøgelser, der viser, at den indre stemmes neurale substrater ligner de substrater, der er aktive, når man taler højt. Derudover kan man også nogle gange måle mikrobevægelser i artikulationsmusklerne (strube, tunge, osv.), mens folk "taler" med den indre stemme. For det andet er der spørgsmålet om, hvordan indre og ydre tale hænger sammen. Det virker plausibelt, at indre og ydre tale udvikler sig i tandem, og at indre tale generelt opleves som en svækket form for udtalt tale (mere komprimeret, svagere og mindre effektiv til f.eks. at understøtte hukommelsen). Funktionerne af indre og ydre tale er sandsynligvis kontinuerlige. For det tredje må vi overveje, om indre tale bedst karakteriseres som indre *tale* eller indre *hørelse*. Disse to kan adskilles ved hjælp af *Descriptive Experience Sampling*-forskning (deskriptive oplevelsesstikprøver), hvor deltagerne bærer en bipper og er trænet i at beskrive deres oplevelse på tilfældige tidspunkter uden forudfattede forestillinger om, hvad deres oplevelse skal indeholde. Indre tale og indre hørelse ser også ud til at have forskellige neurale substrater. Indre tale kan ikke reduceres til hverken indre hørelse (auditiv forestillingsevne) eller

indre tale, selvom sidstnævnte er mere plausibelt, da tale altid ledsages af hørelse (når man taler, hører man altid sig selv). Det omvendte er ikke tilfældet. For det fjerde er der megen debat om, hvorvidt indre tale har alle de samme artikulatoriske og fonologiske træk som ydre tale, eller om indre tale er abstraheret væk fra talens auditive-motoriske træk (nogle gange til sådan en yderlighed, at indre tale ikke har nogle lyde og kan karakterises som "at tænke i rene betydninger"). Nogle teoretikere ser indre tale som værende i det væsentlige motorisk simulation og derfor fuldt specificeret med hensyn til fonologi og artikulatoriske bevægelser. Andre ser indre tale som værende abstraheret væk fra sådanne specifikationer, men hvor den indre stemme stadig bevarer andre sproglige egenskaber såsom hierarkisk struktur, semantik, tempo og rytme. Forskningen indtil videre tyder på, at graden af abstrakthed af indre tale afhænger af både individuelle forskelle og situationelle krav.

I introduktionen til afhandlingen skitserer jeg også to teoretiske perspektiver på den indre tales funktioner, som tjente som udgangspunkt for mit ph.d.-arbejde: fonologisk loop-perspektivet og det Vygotskianske selvreguleringsperspektiv. Ifølge fonologisk loop-perspektivet er indre tale identisk med det fonologiske loop eller den verbale arbejdshukommelse, en hypotetisk komponent i arbejdshukommelsen, hvor fonologisk information genopfriskes, indtil den ikke længere umiddelbart er nødvendig (deraf "arbejdshukommelse"). Der er væsentlig evidens for en sammenhæng mellem det fonologiske loop og den indre tale, selvom indre tale sandsynligvis har en bredere vifte af funktioner end det fonologiske loop. Ifølge det Vygotskianske selvreguleringsperspektiv har indre tale sine rødder i omsorgspersoners regulerende, børnerettede tale, som gradvist internaliseres gennem barnets udvikling og bruges hos voksne primært til selvregulering. Der er også betydelig evidens for dette perspektiv, for eksempel da de fleste børn taler højt for sig selv, og det ser ud til at hjælpe dem med at løse problemer.

Som en del af min kandidatuddannelse (integreret med min ph.d.) gennemførte jeg en undersøgelse af, hvordan atleter taler med sig selv. I den første del af denne undersøgelse brugte vi maskinlæringsmetoder til at forudsige sportstype (badminton eller løb) ud fra spørgeskemadata vedrørende indre tale under sport. I den anden del af undersøgelsen brugte vi igen maskinlæringsmetoder til at forudsige personlige rekordmaratontider ud fra samme spørgeskemadata. Selve kandidatafhandlingen fokuserede også på at kombinere det Vygotskianske selvreguleringsperspektiv med litteraturen om selvrettet tale i sport, og dele af nærværende ph.d.-afhandling bygger på dette tidligere arbejde. Mit fokus på den indre stemmes rolle i fysisk og mental udholdenhed er også delvist inspireret af den sportspsykologiske litteratur.

I mit ph.d.-arbejde kombinerer jeg det Vygotskianske perspektiv og fonologisk loop-perspektivet ved at bruge metoder fra fonologisk loop-forskning til at teste Vygotskianske hypoteser. Specifikt benytter jeg mig af en interferensmetode, hvor vi forsøger at forhindre folk i at bruge deres indre tale til at løse en given opgave ved at bede dem om at udføre en sekundær, simultan verbal opgave (verbal interferens). Logikken bag denne metode er ligetil: Hvis den indre stemme normalt hjælper folk med at løse en given opgave, så burde de være dårligere til at løse den opgave, når deres verbale ressourcer er optaget. Bortset fra interferens med sådanne simultanopgaver fokuserer jeg på oplevelsesstikprøver og individuelle forskelle som metoder til at undersøge indre tale. Ved hjælp af oplevelsesstikprøver beder vi folk på tilfældige tidspunkter rapportere, om de oplever indre tale, og hvilke egenskaber deres indre tale har (indhold, positivt eller negativt, om fortiden eller fremtiden osv.) Dette giver mere fintfølende og pålidelige rapporter end standard retrospektive, generelle spørgeskemaer. Tilgangen, der fokuserer på individuelle forskelle, lader os undersøge, om rapporterede forskelle i oplevelsen af den indre stemme er forbundet med målbare forskelle i adfærd. Dette kan også kaste lys over, hvad indre tale bruges til og ikke bruges til.

I **Artikel I** præsenterer jeg en systematisk gennemgang af studier af verbal interferens. Vi gennemgik 101 undersøgelser og fandt 11 primære opgavekategorier. Disse var kategorisering (simpel og kompleks), hukommelse, hovedregning, motorisk kontrol, ræsonnement (verbale materialer og non-verbale materialer), opgaveskift, social kognition, visuel forandring og visuospatial integration og orientering. Vi fandt også fire forskellige slags interferensopgaver (artikulatorisk undertrykkelse, hukommelsesopgaver, verbal kopiering og verbale bedømmelsesopgaver). Vi fandt evidens for, at den indre tale var involveret i opgaver, hvor deltagerne skulle huske skiftende opgaveregler, holde tal i tankerne for at udføre hovedregning og foretage kategoriske vurderinger. Vi fandt dog ikke overbevisende evidens for involvering af den indre stemme i ræsonnement med non-verbale materialer, social kognition og visuospatial integration og orientering. I artiklen diskuterer vi fordele og ulemper ved verbal interferens som en tilgang til at studere den indre tales rolle i kognition.

Artikel II indeholder to adfærdseksperimenter, hvor deltagerne udførte spurter på en motionscykel, mens de var under den form for verbal interferens, der er beskrevet i Artikel I. For at kontrollere for effekten af at gøre to ting på samme tid sammenlignede vi verbal interferens med visuospatial interferens. Deltagerne var dårligere til at presse sig selv til at cykle hurtigere, når de blev distraheret fra at tale til sig selv – også sammenlignet med generelt at være distraheret. Vores fortolkning af dette var, at folk normalt har gavn af at tale med sig selv for at forhindre sig selv i at stoppe under udholdenhedsopgaver. Hvor meget deltagerne var påvirket af interferens, så ikke ud til at afhænge af, hverken hvor meget de rapporterede, at de normalt taler til sig selv, eller om de rapporterede, at det at tale med sig selv normalt hjælper deres præstation.

I **Artikel III** kombinerede vi forskning i *mind-wandering* (som har testet de skadelige virkninger af opmærksomhedsdrift væk fra en tildelt opgave) og oplevelsesstikprøver for at teste, om folk har nytte af at tale med sig selv, når en opgave er meget kedelig og ikke kræver andet end deres opmærksomhed. Deltagerne så på en tom skærm i flere minutter ad gangen og skulle reagere på sporadiske stimuli meget hurtigt. Efter hver prompt besvarede deltagerne spørgsmål om deres oplevelse i øjeblikket før prompten (hvilket format deres oplevelse var i, om deres tanker handlede om opgaven osv.) Vi fandt ud af, at deltagerne reagerede hurtigere og med mindre varierede reaktionstid på sjældent forekommende prompter, når de talte med sig selv om opgaven. Ligesom i Artikel II tolker vi dette som, at deltagerne er i stand til at bruge indre tale til at forhindre deres tanker i at vandre, dvs. at give op under en mental udholdenhedsopgave.

I **Artikel IV** præsenterer jeg en undersøgelse af mennesker, der oplever lidt eller ingen indre tale, og hvilke adfærdsmæssige konsekvenser dette kan have. Sådanne deltagere (med "anendophasia", som vi kalder det) klarede sig mindre godt i opgaver, der specifikt krævede fonologisk arbejdshukommelse (rimbedømmelse og øjeblikkelig seriel genkaldelse af ord), men lige så godt som sammenligningsgruppen med mere indre tale i opgaver, der krævede visuelle vurderinger og opgaveskift mellem simple additions- og subtraktionsproblemer. Dette indikerer, at selvrapporteret tilbøjelighed til at opleve og bruge indre tale er forbundet med især uudtalt lydsammenligning og verbal arbejdshukommelse. Det er fortsat et åbent spørgsmål, om personer med anendophasia virkelig ikke oplever nogen indre tale overhovedet eller kan siges at opleve elementer af indre tale, f.eks. konceptuel struktur uden lydlige ord (dvs. stærkt abstraheret væk fra auditive og artikulatoriske processer). Dette må udforskes i fremtidige undersøgelser.

Det arbejde, der præsenteres i de fire artikler, har vigtige implikationer for, hvordan vi forstår indre tale. Artiklerne understreger den tætte forbindelse mellem indre tale og verbal arbejdshukommelse, hjælper med at adskille de specifikke mekanismer bag den indre stemmes selvregulerende funktioner og kaster nyt lys over individuelle forskelle i oplevelsen af den indre stemme og dens forhold til adfærd.

Acknowledgments

I am the kind of person who gets nostalgic before-the-fact, so naturally, I have been drafting these acknowledgments since 2021. It's a well-known cliché that a PhD is more a marathon than a sprint but it's also just as much a team effort as it is an individual endeavour. This section is dedicated to the people who made it possible, in big ways as well as small.

Thanks is due to the gang in building 1485 for keeping the days bright and the coffee flowing. When life gives you a global pandemic, you better hope your support system is strong enough to last through Zoom coffee breaks and virtual Cards Against Humanity – and mine was. When not quarantined, I've looked forward to coming into work every day, to air my latest Brilliant Idea (immediately decimated), make you laugh at how long I spent fixing the dishwasher over the weekend, to vent about my latest desk rejection. From you, I have learned that failure and disappointment are inevitable (but not disastrous) parts of academic life but also that so are curiosity, playfulness, and tenacity. In actual fact, we talked more about politics than about research. It will be years before I can repay the debt I owe you for piloting my at times pretty gruelling experiments. We are friends as well as colleagues, and that is a rare privilege. Special thanks is due to my dearest office mate, Line Elgaard Kruse Danielsen, for daily supply of cake, laughs, and empathy.

Neighbouring building 1485 is of course 1483 which houses the wonderful Interacting Minds Centre which has provided me with an endless supply of interesting new perspectives from people from all walks of academia. I have loved the IMC breakfast and writing café but I also owe thanks for a very practical reason: Without the IMC seed funding, I would not have been able to carry out the experiments that are the cornerstones of my PhD or support the strong international collaborations I have.

As much as I have loved working in Aarhus, I also had a lot of fun and mind-boggling experiences visiting other labs around the world. In the spring of 2022, I was a visiting researcher at Gary Lupyan's lab at the University of Wisconsin-Madison. To my immense joy, I there found people who were interested in *exactly the same problems* as me. My stay there would not have been the joyous experience that it was without Meg Meyer, my random AirBnB host who ended up more or less adopting me. In the autumn of 2022, I spent a month as a visiting researcher in beautiful Rome with Anna Borghi's lab, generously supported by the Danish Institute. In both Madison and Rome, I forged friendships and collaborations that I hope will last me years.

As is the case for most people, my PhD has been a steep learning curve. I have written thousands of lines of code, quite substantial numbers of which should never see light of day. The COVID-19 pandemic forced me to pivot to online experiments which would not have been possible without assistance from Malte Lau Petersen, Kenny Smith, and Alisdair Tullo. Despite and between lockdowns, I managed to squeeze in two behavioural experiments at COBE lab for which I should thank Dan Mønster, Trine Fischer Enig, Lasse Lui Frandsen, Marc Hye-Knudsen, and Solveig and Kim Topp. Without their help with recruitment, equipment, and mood upkeep, my PhD would be in a lot worse state right now than it is. This is also the case for the generous assistance from my more or less official co-supervisors and co-authors on the PhD articles: Mark Schram Christensen, Gary Lupyan, and Joshua Charles Skewes.

When the academic woes became overwhelming, I could always turn to the women at Aarhus Rugby Club. I would like to thank you for giving me a few hours every week to think about nothing but mud, grass, spin passes, low tackles, and where to run at exactly the right time. The fun we've had (both with and without beer) has been fully worth the sprains (shoulders, ankle, thumb), concussion, and bent sternum.

My lovely and powerful friends: Klara, Line, Emma, Sara, Fie, Jose, Patryk, Johan, Julie, Anne, Ane, Saga, Peter, Joan, Josie, Letty, Emily, Jacob, Ben. Family and as close to family as you can get. I've never been afraid of falling because I knew you would catch me. I thank you for listening to me vent, for encouraging me when I felt down and directionless, for reading and commenting on bits and pieces, for saying, 'I knew I'd be in the acknowledgments, the question is which thing you're thanking me for'.

My supervisor, Mikkel Wallentin, for picking me up after every failure and celebrating every triumph with me. For taking a chance on me back when you had no idea who I was, for letting me (… actively encouraging me to…) do whatever I wanted, for representing the unpopular opinion in lunchtime arguments, for that two-hour long walk in Risskov when I was nearing the end of my tether, for all the chats both in my head (privileged information of all of Denmark after my Go' Morgen Danmark performance) and in real life, for being my supervisor in life as well as in academia. I promise I'll never buy a house with a thatched roof.

I would not have made it as far as I have without the support of my mum. Thank you for teaching me not to take myself too seriously and to always believe that others have good intentions, both of which attitudes make life in academia much better.

And finally, thank you, Chris, for too many things to count. For correcting my commas, proofreading my high stakes emails, loving Chomsky as much as I do, going on adventures with me, being my home wherever we find ourselves. In every dimension, I'm enjoying doing laundry and taxes with you.

1. INTRODUCTION

Most of us have the experience of talking to ourselves in our heads. This inner speech can take the form of self-encouragement, planning, rehearsing, scolding, practicing speeches and conversations, and a host of others (McCarthy-Jones & Fernyhough, 2011; Morin et al., 2011, 2018; Ren et al., 2016). Given folk intuitions of how prevalent the phenomenon is, it makes sense for a scientist of the mind to ask: How frequent is inner speech actually? What, if anything, is it for? When does it help and when does it hinder? These exact questions are the focus of the present thesis.

In this introductory chapter, I will delve into theoretical perspectives on what inner speech *is* and what it is *for*. It can be thought of as simply a kind of speech rehearsal or verbal working memory (Baddeley & Hitch, 2019; Baddeley & Lewis, 1981; Baddeley & Wilson, 1985) or a biproduct of a perpetually active language production and comprehension system (Pickering & Garrod, 2013). Or as some people see it: indistinguishable from thought itself. We sometimes experience it as helpful and motivational, telling ourselves 'you can do it!'; but if it gets out of control, it can also be detrimental to mental health, as seen in anxiety (McCarthy-Jones & Fernyhough, 2011), depression (Hollon & Kendall, 1980), and auditory-verbal hallucinations (Agnati et al., 2012). It can be difficult to distinguish genuine self-report about how inner speech is experienced from folk psychological notions of what inner speech is, and people are indeed often wrong about their own experience (Hurlburt et al., 2022; Hurlburt & Heavey, 2015). Inner speech research takes place at a busy intersection between many different disciplines (phenomenology, philosophy of mind, psycholinguistics, developmental psychology, cognitive neuroscience, to name a few). Having outlined relevant theoretical perspectives, I will discuss how my own work addresses the functions of inner speech, as well as my choice of methods and their benefits and limitations. There are four articles included in the present thesis, and these four are outlined within the thesis itself. Full versions can be found in the supplemental materials, and all data, code, and preregistrations (if applicable) are available online on the Open Science Forum or GitHub (see links on title pages of relevant articles). After presenting summaries of the four articles, I will discuss my interpretation of how the articles add to our knowledge about the nature of inner speech and its relationship with behaviour and cognition.

1.1. Inner speech: Preliminary definition and functions

I take as point of departure a popular definition of inner speech formulated by Alderson-Day and Fernyhough (2015b, p. 931) in their comprehensive 2015 review of inner speech phenomenology and functions:

> 'Inner speech can be defined as the subjective experience of language in the absence of overt and audible articulation.'

As the authors point out, this definition is agnostic with regard to who "produces" the language and who "hears" it, whether the "language" necessarily has to have *articulatory* or *phonological* properties, and whether *co*vert and *in*audible articulation is necessary. Others have emphasised the auditory-phonological properties of inner speech (Langland-Hassan, 2018) or specifically de-emphasised such properties (Bermúdez, 2018; Gauker, 2018). It is also still debated whether inner speech necessarily entails both *articulatory imagery* (imagined movements of articulatory muscles) and *auditory imagery* (imagined speech sounds). In subsequent sections, I will explore aspects of this definition when I delve into questions of how we can know that people use inner speech ("subjective experience"), how inner and outer speech are related ("absence of overt and audible articulation"), how inner speaking and inner hearing are related (whether *co*vert and *in*audible articulation is necessary), and how embodied inner speech is (to what extent the "language" has articulatory and phonological properties). Regarding this last point, it is debatable whether inner speech that is experienced as abstracted away from articulatory-motor and phonological-auditory features can still be termed inner *speech*. While I recognise this dissonance, I will nevertheless use "inner speech" to describe thought in a linguistic format (at least including the syntactic structures and semantic categories of natural language) in the present thesis as this term maintains the crucial implication of activity or action that unfolds over time (unlike "internal language" or similar terms which connote a static system).

Just as there are many ideas of what inner speech is, there are also many ideas of what its functions and purposes are. In psychology, two of the most influential accounts have connected inner speech to verbal working memory (Baddeley et al., 2001; Baddeley & Hitch, 1974; Baddeley & Lewis, 1981; Baddeley & Wilson, 1985) and self-regulation in development (Alderson-Day & Fernyhough, 2015b; Cragg & Nation, 2010; Vygotsky, 1962; Vygotsky & Luria, 1994). In the working memory accounts, inner speech is tied to verbal working memory through the phonological loop, a hypothesised dedicated part of working memory where phonological content is rehearsed and thereby kept in working memory (see Figure 1). The phonological loop is often equated with the inner voice and inner ear (Baddeley & Hitch, 2019; Baddeley & Lewis, 1981). One of the primary sources of evidence for the phonological loop comes from dual-task interference studies. In these studies, participants are asked to simultaneously talk out loud or listen to speech to disrupt the functions of the phonological loop (J. D. Larsen & Baddeley, 2003; Saito, 1993). If performance on a primary task is reduced under verbal dual-task interference but not under a control interference task, this is taken as evidence that the phonological loop is involved in the primary task under normal circumstances. According to an influential theory by Russian developmental psychologists Vygotsky and Luria, inner speech is the end point of a developmental trajectory where caregivers' verbal instructions to children is gradually internalised over the first decade of life (Vygotsky, 1962; Vygotsky & Luria, 1994). One of the primary functions of inner speech is thus behavioural self-regulation. The starting point of my PhD work was the idea of combining these two paradigms – inner speech for self-regulation and inner speech as a verbal working memory process – and using dual-task interference methods from the latter to experimentally test hypotheses from the former. However, if we are to make claims about what inner speech is for, then we first need a clearer idea of what it is.

In the following, I will move from questions about the nature of inner speech to questions about its functions and purpose. Before moving onto introducing my own studies, I will discuss in some detail the Vygotskian approach and the phonological loop account.

1.2. How can we know that people use inner speech?

1.2.1. The survey method

The most straightforward way to find out whether people use inner speech is to ask them. There are several different ways of doing this, for example through self-report questionnaires (e.g., asking them how often they say specific phrases to themselves) or through experience sampling (Csikszentmihalyi et al., 1977; Csikszentmihalyi & Larson, 1987; Hormuth, 1986) where people are asked to report what their experience is like at randomly sampled moments. Questionnaire-based measures of inner speech have revealed that most people believe they talk to themselves very frequently and that they use inner speech for self-regulation, motivation, self-reflection, and as a mnemonic aid (Brinthaupt et al., 2009; McCarthy-Jones & Fernyhough, 2011; Morin et al., 2011, 2018; Nedergaard et al., 2021). However, there has been substantial debate about the validity of questionnaire-based approaches (Alderson-Day & Fernyhough, 2015a; Hurlburt et al., 2022; Hurlburt & Heavey, 2015; Uttl et al., 2011). Criticisms are commonly rooted in findings that people frequently misremember or reinterpret their subjective experience (Hurlburt et al., 2013), and that questionnaire-based results differ from experience sampling-based results (Alderson-Day & Fernyhough, 2015a; Hurlburt et al., 2022). Misrepresentation especially happens when reports take place retrospectively, e.g., 'what did you say to yourself during your marathon last month?' (Ptacek et al., 1994; Snelgrove & Havitz, 2010; Wells & Loftus, 2003), or if people are asked to theorise about causes and effects of their subjective experience, e.g., 'do you think it helps you to talk to yourself?' (Berger et al., 2016; Johansson et al., 2005; Petitmengin et al., 2013).

1.2.2. Descriptive Experience Sampling

To counter these self-report weaknesses and to obtain more fine-grained, detailed access to subjective experience, Hurlburt and colleagues developed the Descriptive Experience Sampling method (DES) where participants carry a beeper and have to note down what their subjective experience contains at the exact moments of random beeps – typically six times a day. Shortly after a day of such experience sampling, participants discuss their notes with a DES interviewer. Participants are trained iteratively to report their experience without bias, to "bracket presuppositions" about what their experience may contain. During DES procedures, participants report five distinct – but not mutually exclusive – kinds of experience that DES researchers have identified as inner speaking, inner seeing, sensory awareness, feelings, and unsymbolised thinking. While it is highly labour-intensive, DES has revealed many interesting things about inner speech: people report talking to themselves during around 25 % of sampled moments (Heavey & Hurlburt, 2008), they are quite often initially mistaken about how much or how little time they spend talking to themselves (Hurlburt et al., 2022), people can readily distinguish between the experience of self-produced inner speaking versus hearing a voice that they did not themselves produce (inner hearing) (Hurlburt et al., 2013), and there are large individual differences in how much people experience inner speech (Hurlburt et al., 2013).

Such phenomenological methods also indicate another way that people can tell that their inner speech is indeed speech-like: bilingual people report that they can tell which language they are thinking in (Hammer, 2019; S. F. Larsen et al., 2002; Resnik, 2018), and this has also been found in most instances of inner speech sampled using DES, although it has not been the focus of such research (Hurlburt et al., 2013). The fact that we can tell which natural language inner speech takes place in has been used to argue that inner speech must have phonological properties (Langland-Hassan, 2018). The argument here is that other indicators of a specific natural language – e.g., semantics or syntax – would be less reliable to use to differentiate between languages. Langland-Hassan further argues that inner signing can be recognised as being keyed to a specific language because sign languages also have phonology in the sense that their signs are 'created from a finite list of meaningless elements that combine and recombine' (Sandler, 2012, p. 162). Research on inner signing is sparse (but see MacSweeney et al., 2008; McGuire et al., 1997). In the present thesis, I will focus on covert spoken language – however, note that I expect that my research into inner speech can be applicable and extendable to inner signing as well, apart from aspects of inner speech function and nature that are tied specifically to articulatory and auditory features.

1.2.3. Physiological measures of inner speech

There are also ways of investigating inner speech that rely less heavily on participants' subjective experience. These studies mostly measure people internally verbalising specific words requested by the experimenter rather than spontaneous inner speech (Hurlburt et al., 2016). The method that has been used for the longest time is electromyography (EMG) with electrodes on the lips and tongue (Garrity, 1977; Jacobson, 1930, 1932). With such methods, researchers measure micromovements of the articulators associated with inner speech (throat, tongue, etc.) Recent experiments have found that while it is often possible to measure the presence or absence of inner speech, the measurements may not be fine-grained enough for specific phonetic features (Nalborczyk et al., 2020) or for more spontaneous inner speech, such as rumination (Moffatt et al., 2020; although see Nalborczyk, 2022). Neuroimaging methods have also shown that inner speech is correlated with (attenuated) activity in similar (but not completely identical) areas as outer speech – for example the left inferior frontal gyrus and the inferior parietal lobe (Christoffels et al., 2007; Frings et al., 2006; Hoeft et al., 2007; Lurito et al., 2000; Owen et al., 2004; Paulesu et al., 1993; Poldrack et al., 2001; Pugh et al., 1996). However, such neuroimaging studies have also indicated that inner speech cannot be reduced to outer speech without the motor component (Geva et al., 2011), as activation in both auditory areas and articulatory-motor areas is sometimes absent in inner speech tasks (Hurlburt et al., 2016; McNorgan, 2012). Such studies have highlighted both important commonalities and differences between inner and outer speech.

1.3. How are inner and outer speech related?

We might ask how inner and outer speech are experienced as similar or different for adults, or how they are related in development. Is inner speech a consequence of outer speech (Vygotsky, 1962), or do we learn to speak internally before we learn to speak out loud (Blonskii, 1964)? In this section, I will discuss these different questions in turn.

According to Vygotsky's theory, inner speech comes from gradually internalised speech directed at regulating the child's behaviour (Vygotsky, 1962; Vygotsky & Luria, 1994). That is, caregivers' speech is mimicked in the child's own overt private speech before developing into completely covert speech. Most of Vygotsky's claims about the nature and functions of inner speech were based on observations of children's private speech (because 'the transition between external and inner speech is to be found in the child's egocentric speech'; Vygotsky, 1962, p. 113). From such observations, he concluded that inner speech was condensed (lacking complete syntactic structures and having "personal", idiomatic semantics, i.e., meanings that would only make sense to the person producing them) and dialogic because what is internalised is interpersonal dialogue (Fernyhough, 1996, 2004). One of Vygotsky's main contemporary critics, Blonskii, suggested instead that infants repeat the speech sounds that they hear internally, thus producing inner speech which in turn plays a vital role in learning to produce language (Blonskii, 1964). Indeed, brain areas associated with verbal working memory and inner speech appear to be selectively active already in preverbal infants (Dehaene-Lambertz et al., 2006), and there is evidence that infants can both implicitly name objects, resulting in priming effects (Mani & Plunkett, 2010), and internally produce word forms that they cannot yet produce overtly (Ngon & Peperkamp, 2013). It may be the case that inner and outer speech interact differently over the course of development, i.e., that inner speech at some stages helps practice outer speech and at other stages reproduces self-regulatory statements from external sources.

When researching adult inner speech, we rarely examine participants' overt, private speech (although there are exceptions – see e.g. Macbeth et al., 2022; Thibodeaux & Winsler, 2018, 2022) and instead have to rely on self-report. Several such self-report-based studies have confirmed that inner speech is indeed experienced as more "condensed" than outer speech (Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011). There is also behavioural evidence for condensation from spontaneous covert speech in problem-solving (Korba, 1990) and covert speech production (Anderson, 1982; MacKay, 1981; Marshall & Cartwright, 1978, 1980). However, some Descriptive Experience Sampling findings contradict the Vygotskian characterisation of inner speech; for example, dropped subjects (e.g., 'Want food' instead of 'I want food') only occasionally appear, and there are apparently no more idiomatic, private meanings than there are in overt speech (Hurlburt & Heavey, 2018). A potential explanation for this could be that participants "translate" in honour of the interviewer, despite best efforts to preserve "pristine" inner experience, or perhaps that there is more condensation in overt speech than considered by Vygotsky when interlocutors' perspectives are aligned.

It may not be possible to determine whether inner speech or outer speech comes first in development, and they are likely to develop in tandem. Furthermore, it seems clear that inner speech is experienced as somewhat attenuated in comparison with outer speech and may therefore have different effects – for example, people also benefit more from rehearsing out loud than rehearsing covertly (Dell & Repka, 1992; Keeney et al., 1967). This could partly be due to the fact that when we participate in overt language practices, we not only speak but also listen. The following section addresses whether inner speech includes both speaking and listening.

1.4. Is it inner speaking or inner hearing?

Related to considerations of inner versus outer speech is the question of whether the experience is most properly construed as inner *speaking* or inner *hearing*. Of course, one might ask what the point is of telling these two apart when they usually co-occur in overt speech. The first answer is that the two elements disentangle in overt speech as the difference between talking yourself (both speaking and hearing) and listening to other people talk (only hearing). In inner speech, those two types may be at least partly translated into the difference between inner speech and auditory verbal hallucinations (inner hearing without inner speaking) (Fernyhough, 2004; Perona-Garcelán et al., 2017). Furthermore, it may matter for the design of experimental methods like dual-task interference whether we attempt to interfere with hearing or speaking.

Apart from research on hallucinations, evidence for a distinction between inner speaking and inner hearing comes primarily from neuroimaging research, verbal working memory research, and Descriptive Experience Sampling. In neuroimaging research, Tian and Poeppel (2012) and Tian, Zarate, and Poeppel (2016) for example found that two ways of generating auditory-verbal imagery – through motor simulation and through imagined speech/inner hearing

– were indeed distinct in terms of neural mechanisms. While both activate auditory sensory cortices, the motor simulation route relies on activation of prefrontal and motor areas associated with the articulation of speaking, while the imagined speech/inner hearing route is more like other kinds of sensory imagery (such as visual imagery), derived from previous experiences of sensory events (Wilkinson & Fernyhough, 2018). Lu et al. (2022) also found that listening to speech and imagining speech showed both overlapping activation (bilateral superior temporal gyri and supplementary motor areas) and distinct activation (left inferior frontal gyrus for imagined speech), providing further support for involvement of both auditory and articulatory imagery. With regard to working memory as well, it has been debated whether verbal working memory is best characterised as a phonological loop or an articulatory loop (Baddeley & Hitch, 2019). The question of the extent to which both phonological and articulatory codes exist in inner speech can be approached by using articulatory suppression (interfering with inner articulation by overt repetition of simple syllables) together with homophone and rhyme judgment tasks (Baddeley & Lewis, 1981; Besner et al., 1981). Rhyme judgments are affected by articulatory suppression and likely rely on an articulatory code because the two words to be judged have to be kept in working memory to be manipulated and compared. Homophone judgments, on the other hand, are not affected by articulatory suppression and thus appear solvable through a non-articulatory, auditory code (Baddeley & Lewis, 1981; Besner et al., 1981; Richardson, 1987). DES research supports the distinction between inner speaking and inner hearing: participants are generally able to report whether the inner speech they experience is produced by themselves (inner speaking) or not (inner hearing) (Hurlburt et al., 2013). It should be noted that inner hearing (auditory imagery) can take many forms – music, animal sounds, other people's voices – but it can also be one's own voice, e.g., memory-based as when you recall something embarrassing you have said. Still, as Hurlburt and colleagues say:

> 'In speaking, the words arise from you, are driven by you, emanate from you, proceed away from you; in hearing, the words arise from elsewhere, are outside your control, emanate somewhere else, come toward you.' (Hurlburt et al., 2013, p. 1485).

Hurlburt and colleagues report that their impression is that inner speaking is considerably more frequent than inner hearing.

Despite these apparent phenomenological, neurological, and behavioural differences, there is an argument for not making the distinction that warrants attention: that producing speech overtly is almost always accompanied by hearing speech (when we hear our own voice) whereas hearing speech is almost never accompanied by overtly producing speech, because we tend to be quiet when we listen to others speak. Thus, we could simply think of inner speech as motor-articulatory simulation, and we would expect predictions of auditory consequences to appear in inner speech as a side effect of this motor-articulatory simulation (Tian et al., 2016; Tian & Poeppel, 2012). Indeed, this conjunction of inner speech production and perception fits well with theories of overt speech that emphasise the role of speech production in speech perception (see e.g. Galantucci et al., 2006). In mental imagery research, in contrast, inner speech is often described solely as a kind of auditory imagery (e.g., 'auditory representations and auditory imagery, including the notion of an "inner voice" are thought to be critically important for psychological functioning across a wide range of domains'; Hinwar & Lambert, 2021, p. 1; see also Monzel et al., 2022). This both risks neglecting important motor-articulatory components of inner speech and makes it difficult to form theories of inner speech that can also encompass inner signing. I believe that this characterisation of inner speech (which I will continue to call it for convenience while recognising that it may encompass both inner speaking and inner hearing) as auditory imagery alone neglects a vital action-based component for which there is substantial evidence (Loevenbruck et al., 2018; Nalborczyk et al., 2020; Tian et al., 2016; Tian & Poeppel, 2012). Instead, inner hearing accompanies inner speaking as a sensory consequence or prediction of the simulated action, while the auditory imagery of other people's voices to a greater extent emerges from our real-world roles as listeners in conversations (Tian et al., 2016; Tian & Poeppel, 2012). The view that inner speech is grounded in articulatory-motor simulation is somewhat at odds with another influential view: that inner speech is best understood as abstracted away from auditory and articulatory-motor imagery. I will discuss this tension in the following section.

1.5. How abstracted or embodied is inner speech?

We can think of inner speech as varying on a spectrum from fully linguistically specified and expanded with all the features of spoken-out-loud language to completely abstracted away from overt features of spoken language with only "soundless words" and the semantic and syntactic structures of language remaining. Views at the former end of the spectrum (going back at least to the mid-19th century; Erdmann, 1851; Geiger, 1868) focus on the embodied properties of inner speech. One of the most prominent modern instantiations is the *motor simulation view* according to which inner speech is mental simulation of overt speech without the complete articulatory-motor actions (Grèzes & Decety, 2001; Jeannerod, 1994; Postma & Noordanus, 1996). At the other end of the embodied-abstracted spectrum of perspectives on inner speech is the *abstraction view* which sees inner speech as divorced from bodily experience (Bermúdez, 2018; Gauker, 2018). Here, inner speech plays a role in cognition not necessarily because its functions are continuous with those of overt speech but because it provides the categorical, hierarchical, compositional, self-referential structures of language. This view is related to classical symbol-processing theories of cognitive architecture (Fodor, 1975; Fodor & Pylyshyn, 1988) which have recently regained popularity (Mandelbaum et al., 2022). In the following, I will consider evidence for the motor simulation view and for the abstraction view as well as a potential reconciliation.

As previously discussed, there is considerable evidence that inner speech has motor components as measured both through micromovements of articulatory muscles (Livesay et al., 1996; McGuigan & Dollins, 1989; Moffatt et al., 2020; Nalborczyk et al., 2020; Sokolov, 1968) and neural activations of motor and premotor cortices during inner speech tasks (Basho et al., 2007; Huang et al., 2002; McGuire et al., 1996). However, the most extreme version which states that articulatory movement is *necessary* for inner speech is untenable – see e.g., Smith, Brown, Toman, & Googman (1947) who paralysed the mouth and tongue and still found intact inner speech. In addition to physiological and neural correlates, behavioural measures such as sentence rehearsal also correlate in covert and overt speech (MacKay, 1981, 1992) and reading (Abramson & Goldinger, 1997). However, inner speech can be faster than outer speech (MacKay, 1981, 1992), and speaking out loud does not necessarily disrupt the ability to simultaneously monitor inner speech (Wheeldon & Levelt, 1995). It is thus more a question of *how* embodied and concrete inner speech is rather than *whether* it is abstracted or embodied. To get a more precise idea of the degree of phonological and lexical specification in inner speech, we can also ask people to say tongue twisters in their heads (e.g., 'she sells seashells by the seashore') and report their mistakes (Dell & Repka, 1992; Postma & Noordanus, 1996). This method has for example shown that the lexical bias effect persists (i.e., mistakes that make real words are more likely than mistakes that make nonsense words) but that only overt speech errors are affected by the similarity of articulated phonemes (e.g., "reef" is more likely to become "leaf" than "beef" because both /r/ and /l/ are voiced approximants) (Oppenheim & Dell, 2010; but see Corley et al., 2011). This suggests that inner speech includes lexical specification but perhaps not (always) phonological specification.

One way of resolving an abstracted conception of inner speech with an embodied one is to argue that there may be different levels of auditory and motor abstraction in inner speech (Fernyhough, 2004; Grandchamp et al., 2019; Oppenheim & Dell, 2010). This somewhat parallels influential theories of language production where utterance planning and execution pass through conceptual preparation and grammatical encoding before morphophonological and phonetic encoding and ultimately articulation (Levelt, 1989, 1999). Going from more abstracted to more embodied, the first level might correspond to Vygotsky's description of the most condensed form of inner speech as 'thinking in pure meanings' (Vygotsky, 1962). At this level, the phonological features of overt speech are thought to be lost (Fernyhough, 2004). Thinking about inner speech as thus abstracted away from auditory-articulatory processes may explain why participants in Descriptive Experience Sampling studies sometimes report "wordless" inner speaking (Hurlburt et al., 2013). In such episodes, participants experience the agency of producing language as well as the pace, rhythm, and sense of unfolding over time – but without the experience of hearing or speaking words (Hurlburt & Heavey, 2006). "Wordless" inner speaking can be likened to a series of tip-of-the-tongue states. To what extent people experience their inner speech as expanded and linguistically specified may depend on where in the "production simulation process" the speech was abandoned. This in turn is likely to vary across both individuals and situations. For example, Fernyhough (2004) suggests that inner speech may become more like overt speech as a function of rising cognitive demands (or cognitive disruption; see also Brinthaupt, 2019). The condensed-expanded dimension of inner speech is both relevant to the functions of inner speech (e.g., one function could be fine-grained speech planning and error prevention which would need highly specified inner speech) and how we study it (e.g., it is not clear that articulatory suppression should have an effect on inner speech functions if inner speech is divorced from perceptual-motor processes).

1.6. What is inner speech "for"?

My PhD work has been centred around the question of what inner speech is for rather than what inner speech is, although these questions are of course intertwined. Below, I will outline two of the dominant views of inner speech functions – the Vygotskian perspective and the phonological loop perspective. These are the most relevant ones in the present context, as my work focuses on self-regulatory functions of inner speech while examining and using the dual-task interference method to a large extent.[1]

1.6.1. The Vygotskian perspective

The Vygotskian perspective on inner speech (Vygotsky, 1962; Vygotsky & Luria, 1994) has been highly influential during the 20th century and continues to be so. According to this view, inner speech as adults experience it is the result of a developmental trajectory of internalisation of caregivers' instructions to children. During this process, child-generated private (overt, self-directed) speech in the absence of caregivers is transformed into completely covert speech. Caregivers tell the child what to do and what not to do and help direct their attention towards relevant aspects of the environment and solve problems. These are the functions that are hypothesised to be present in self-regulatory inner speech as well which are thought to enable deliberate planning and organising of actions in pursuit of specific goals (Vygotsky, 1962; Vygotsky & Luria, 1994). During the internalisation process from private speech to inner speech, both the semantics and syntax of private speech may be transformed (e.g. with condensation, as previously discussed), but the functions are hypothesised to remain largely the same (Winsler & Naglieri, 2003).

There is substantial evidence for the Vygotskian view of inner speech as being for self-regulation. First, private speech is extremely prevalent in children with one study finding that 93

[1] Other perspectives include (but are not limited to) those with a dual-process focus where spontaneous and deliberate self-talk respectively serve as vehicles for System 1 thinking (intuitive, fast, effortless) and System 2 thinking (slow, effortful, demanding) (Van Raalte et al., 2016), accounts that focus on inner speech as practice for communication (Feigenbaum, 2009), and the predictive processing framework (Grandchamp et al., 2019; Loevenbruck et al., 2018) according to which inner speech consists of predictions of speech simulation outcomes.

% of three-year-olds engaged in it (Winsler et al., 2003). Second, most of this private speech concerns control of the child's own actions (Berk & Garvin, 1984). Third, private speech is more prevalent when the child is trying to solve a difficult problem alone, and more private speech is associated with better task mastery (Berk, 1994; Bivens & Berk, 1990). However, we cannot necessarily extrapolate from private speech in children to inner speech in adults, and it may also be the case that children who engage more in private speech are simply also more cognitively developed, and so the connection between private speech and task performance is not necessarily causal. Nevertheless, the finding that more private speech is associated with better task performance is remarkable. That caregivers' child-directed speech should be successful at regulating behaviour and improving how children solve both physical and social tasks is not surprising. However, the fact that children can speak to themselves to control their own attention and actions is of great theoretical interest due to studies indicating similar facilitative effects in adults (Emerson & Miyake, 2003; Kray et al., 2008; Liefooghe et al., 2005; Rao & Baddeley, 2013).

Inner speech appears to continue to play an important, facilitative role in adult cognition, and the use of verbal strategies to solve tasks increases during adolescence (Chevalier & Blaye, 2009; Kray et al., 2008). In many studies on inner speech in adult cognition, there is a similar focus to the one found in developmental studies on cognitive flexibility and task switching (Baddeley et al., 2001; Emerson & Miyake, 2003; Goschke, 2000; Liefooghe et al., 2005; Miyake et al., 2004; Saeki et al., 2013). As in the developmental reports, the studies in adults have generally found support for the idea that inner speech aids task retrieval, task rule maintenance, and task order maintenance. There is evidence for a similar self-regulatory role of inner speech in sport psychology (Hatzigeorgiadis et al., 2011; Tod et al., 2011), although this literature remains relatively unconnected to general cognitive psychology theories. When playing sports, inner speech may be recruited for drawing attention to details of motor movements (Galanis et al., 2021; Hatzigeorgiadis et al., 2004; Hatzigeorgiadis & Galanis, 2017), remembering and maintaining rules, and motivation in endurance sports such as running, swimming, and cycling (Barwood et al., 2015; McCormick et al., 2018; Stets et al., 2020). Sport psychology also makes contact with Vygotskian self-regulation in a focus on the discursive nature of inner speech where inner speech does not stem from a single entity (the self) but rather arises from an internal "conversation" between voices that have been internalised during development throughout childhood and beyond (the voices of societal narratives, personal relationships, coaches, etc.) (Fernyhough, 1996; Larrain & Haye, 2012; Van Raalte et al., 2016).

Despite the Vygotskian perspective's success, there are also aspects that have been criticised. First, Vygotsky believed that self-regulation was the primary function of private speech. Later empirical research has, however, found that self-directed talk has multiple functions (pretence, practice for social encounters, language practice, etc.; Berk, 1992). Second, the idea that inner speech comes from private speech does not fit well with the finding that the development of the phonological loop predates private speech (Hitch et al., 1991; Mani & Plunkett, 2010; Ngon & Peperkamp, 2013; Perrone-Bertolotti et al., 2014). Having a phonological loop implies being able to rehearse verbal information internally and therefore potentially engaging in covert, self-directed speech. Third, some of Vygotsky's specific ideas about syntactic condensation and idiosyncratic semantics may not hold up to scrutiny with Descriptive Experience Sampling, as previously discussed, but they may be relevant for the discussion of the embodied versus abstracted nature of inner speech. More modern versions of the Vygotskian perspective have to a large extent focused on the self-regulatory aspects of inner speech (Al-Namlah et al., 2006; Cragg & Nation, 2010; Fernyhough, 2004) whereas the condensation dimension has been connected to the abstracted-embodied spectrum (Alderson-Day & Fernyhough, 2014; Fernyhough, 2004), as discussed in section 1.5. 'How abstracted or embodied is inner speech?' above.

1.6.2. The phonological loop perspective

Inner speech has strong theoretical ties to verbal working memory, and several of the methods used to investigate verbal working memory have been used to investigate inner speech as well. Working memory refers to the online retention of information that is necessary to solve immediate tasks. The influential multicomponent model of working memory (Baddeley & Hitch, 1974) posits that we have dissociable verbal working memory (phonological loop and store) and visuospatial working memory (visuospatial sketchpad), and that these can be occupied or interfered with separately (see Figure 1 for a sketch of the multicomponent working memory model). It is hypothesised that we use the phonological loop for continuously refreshing phonological information to keep it "in mind". In the initial conceptualisation of the sketchpad and phonological loop, they were thought of as simple storage and rehearsal components under the control of the central executive, the component that delegates and allocates resources. It has become increasingly apparent, however, that the simple subsystems can also be used for action control (Baddeley et al., 2001; Emerson & Miyake, 2003; Miyake et al., 2004).

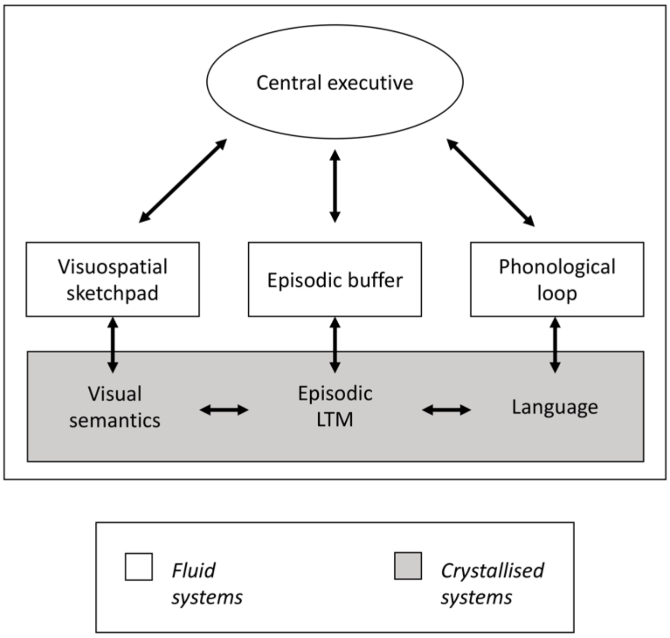


Figure 1. *The multicomponent model of working memory (LTM = long-term memory). Adapted from Baddeley (2012).*

As can be seen in Figure 1, the working memory model also includes a multimodal "episodic buffer" that can bind features together and access long-term memory (Baddeley, 2000, 2007) as well as links between working memory components (fluid systems) and long-term memory components (crystallised systems). For our present purposes, the phonological loop is the most important as it is often treated as equivalent to inner speech:

> '[the phonological loop is an] articulatory rehearsal process that is analogous to subvocal speech' (Baddeley, 2003, p. 830).

Depending on the exact model instantiation, verbal working memory sometimes includes both a phonological *loop* where information is rehearsed and refreshed and a phonological *store* which can store verbal information passively for a few seconds (Baddeley & Hitch, 2019; Baddeley & Larsen, 2007; Jones et al., 2004, 2007).

Research on the phonological loop has mostly focused on empirically testing whether it is indeed dedicated to phonological/articulatory information, and whether there is both a phonological loop and a phonological store (Baddeley & Larsen, 2007; Jones et al., 2004, 2007). Evidence for a dedicated verbal working memory usually comes from immediate serial recall of digits, letters, or words – the features of these materials to-be-remembered and how well they are recalled are used as evidence for a phonological rehearsal mechanism. For example, phonologically similar letters (in English) such as V, B, G, T, P, C are less well recalled than phonologically dissimilar letters (W, X, K, R, Y, H) (the *phonological similarity effect*) (Conrad, 1964; Conrad & Hull, 1964). Similar effects are observed with immediate word recall. When long-term memory is instead involved, (dis)similarities in meaning become more important than similarities in sound (Baddeley, 1966a, 1966b). Corroborating evidence comes from experiments using dual-task interference where participants attempt to solve a task (such as remembering a string of letters) while a simultaneous secondary task occupies their phonological loop. Such experiments find, for example, that the phonological similarity effect disappears with articulatory suppression (J. D. Larsen & Baddeley, 2003; Logie et al., 1996; Neath et al., 2003).

The distinction between a phonological loop and a phonological store does not, perhaps, matter so much for inner speech research (but see Jones et al., 2004, 2007, for discussion). What does matter, however, is whether there is a dedicated component of working memory for verbal material. This matters for inner speech research both because it speaks to the embodied or abstracted nature of inner speech and because the dual-task method to some extent relies on this assumption. First, if inner speech is identical with the phonological loop and if the phonological loop always has auditory-phonological representation, then it is difficult to imagine where there would be a place for inner speech that is abstracted away from auditory-phonological representation. Second, if dual-task interference (such as articulatory suppression) disrupts the language-specific processes of the phonological loop, then such interference should also disrupt other internal uses of language (problem-solving, self-cuing, planning, etc.) This is perhaps the most important link between working memory research and inner speech research – it gives us a way of (albeit somewhat crudely) conducting empirical tests of the functions of inner speech. We discuss the advantages and disadvantages of using verbal interference to test the role of inner speech in behaviour and cognition at length in Article I of the present thesis.

The phonological loop perspective has been highly useful and provided detailed experimental paradigms to test inner speech. However, we should not completely reduce inner speech to verbal working memory, even if they are related as appears to be the case. The phonological loop may be the path that verbal information travels on, but what we use this verbal information for likely goes beyond maintaining and storing verbal information, e.g., for self-regulation, self-instruction or simulating conversations (Baddeley, 2012).

1.6.3. How compatible are these accounts?

There is much to gain from combining Vygotskian self-regulation and the verbal working memory perspective. The Vygotskian ideas provide a robust, well-tested explanation for the ontogenetic origin of inner speech and for some of the functions in self-regulation that it serves in adults. I have used this perspective as a starting point when testing the self-regulatory functions of inner speech (see Articles II and III). The phonological loop perspective provides perhaps the most rigorous, fine-grained experimental methods and clear evidence that inner speech does in fact in most cases have auditory-articulatory representation. I have explored the limitations and potential of such methods to investigate inner speech functions in two of my articles as well (Articles I and II).

The connection between the Vygotskian self-regulatory functions of inner speech and verbal working memory functions becomes especially clear when we discuss dual-task interference experiments. As previously discussed, verbal interference is designed to occupy the phonological loop, and dual-task experiments testing self-regulatory functions of inner speech rely on the assumption that such self-regulation indeed takes place using the phonological loop (e.g., Baddeley et al., 2001). Concretely, the phonological loop could for example be used to repeat task instructions or rules, and there is evidence that it is used in such ways during task switching tasks where participants have to switch quickly and flexibly between different rules (see Article I). This is exactly the kind of self-regulation that the Vygotskian perspective on inner speech predicts. However, there are also aspects of the two perspectives which are less compatible. For example, it is a central part of the Vygotskian conception of inner speech that it should be condensed in terms of both its syntax, phonology, and semantics (Vygotsky, 1962). It is not clear how inner speech can be held as equivalent to the phonological loop if inner speech does not (necessarily) have fully specified phonological representation. The phonological loop account assumes that inner speech is fully specified and needs both articulatory and auditory imagery – indeed most of the experiments described in section 1.6.2. 'Phonological loop perspective', such as the ones exploring phonological similarity effects, target quite fine-grained phonological differences.

1.7. Methodologies and motivations

To study inner speech, we should first verify that inner speech is in fact present. While one commonly used approach is to rely on self-report, there are, as previously discussed, also ways of verifying the presence or absence of inner speech that are less reliant on reports of subjective experience (silent rhyme judgments, silent tongue twisters, neuroimaging, verbal interference, etc.) I decided to focus mostly on verbal interference and self-report because I found that both the limitations and potential of verbal interference were underexplored, and I was interested in conditions under which reports about inner speech experience can be valid and reliable. Article IV also includes some objective measures of inner speech instantiated by investigating silent rhyme judgments and verbal working memory capacity.

With my investigations, I have attempted to approach the question of what inner speech is used for from several angles. Philosophical articles about the role of language in cognition (e.g., Carruthers, 2002; Clark, 1998; Frankish, 2018) inspired me to think critically about the dual-task interference method and its limitations as a method as well as areas where its potential was underexplored. Such ideas became the systematic review of verbal interference presented in Article I of the present thesis. To explore the verbal interference method myself, I conducted two behavioural experiments where I used the method to examine the role of inner speech in the control of physical endurance (presented in Article II). Article II was a conjunction between sport psychology research and the dual-task interference method borrowed from working memory research. I first explored sport psychology as a testing ground for theories of inner speech functions in my master's thesis, which was published as the paper 'Valence, Form, and Content of Self-Talk Predict Sport Type and Level of Performance' (Nedergaard et al., 2021). In the paper, we found that self-reported self-talk was predictive of whether participants were runners or badminton players (Study 1) as well as how proficient they were at running marathons (Study 2). While these were intriguing correlations, working on the paper sparked my interest in more direct ways of measuring the effects of inner speech on behaviour, such as dual-task interference (resulting in Article II). I also became increasingly aware of the limitations of questionnaire studies of inner speech and interested in experience sampling. I chose to leverage the potential of experience sampling through an online study of the role of inner speech in sustained attention (Article III). This was both related to the idea that inner speech is recruited for self-control and endurance in sport (and thus also perhaps "mental" endurance) and the idea that we may gain insights into inner speech from experience sampling research.

Some of the most debated questions in inner speech research concern the embodied or abstracted nature of inner speech as well as how this may vary across situations and individuals (Fernyhough, 2004; Grandchamp et al., 2019; Langland-Hassan, 2018; Oppenheim & Dell, 2010). Variations in inner speech experience across individuals is important to study because they can provide insights into verbally mediated strategies, the embodied or abstracted nature of inner speech, and what cognitive processes (if any) expanded inner speech is *necessary* for. Therefore, I also present a study on what we call "anendophasia", the phenomenon of experiencing little to no inner speech, and whether this has any consequences for behaviour (Article IV).

2. THE ARTICLES

2.1. Article I: Verbal interference systematic review

2.1.1. Background

Verbal interference has, as mentioned, been employed in a wide range of contexts to test the putative role of language in different cognitive tasks. The method originates in studies of verbal working memory where it was used to provide evidence for the existence of a phonological loop as a separable, dedicated component of working memory (Baddeley & Hitch, 2019; Baddeley & Larsen, 2007). As described above, the logic behind dual-task interference is straightforward: if the primary task (e.g., detecting colour differences) depends on verbal resources, then performance should be worse on the primary task if verbal resources are occupied with a concurrent verbal task compared with a concurrent non-verbal control task (see Figure 2). To take a straightforward example, if you are repeating a phone number in your head while also repeating "the" out loud, your memory of the phone number will be poorer than if you had been tapping your foot while repeating the phone number.

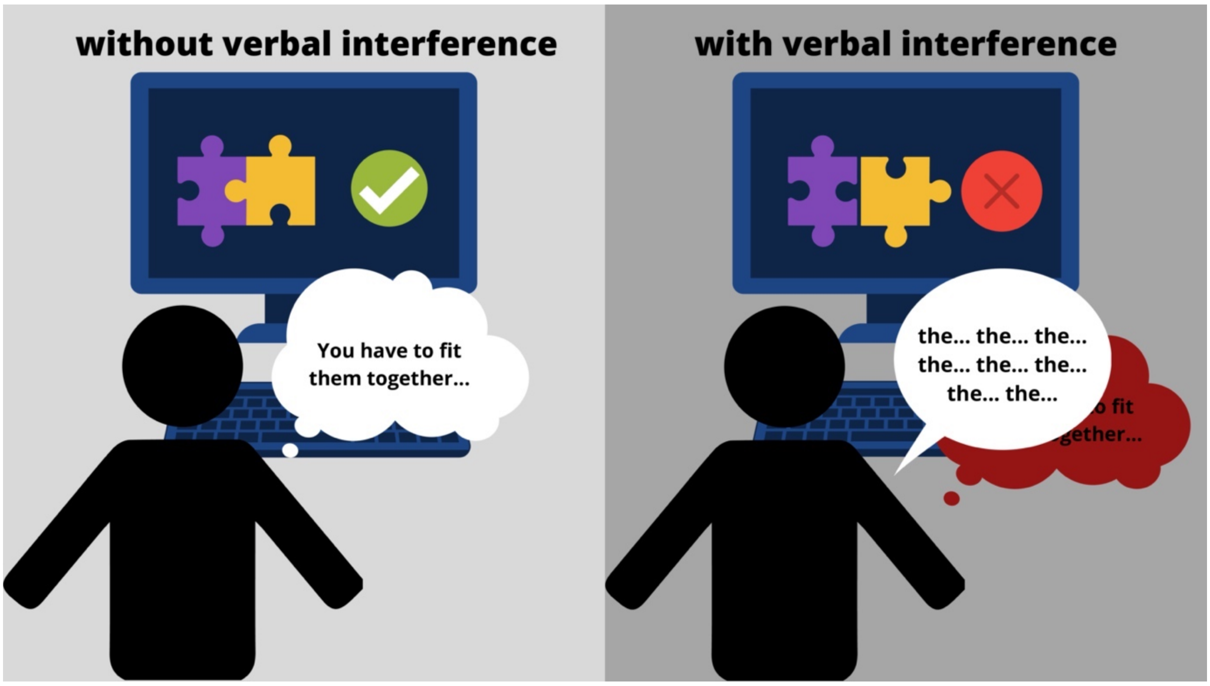


Figure 2. *A sketch of the logic behind verbal interference studies. In this case, the primary task would be fitting the pieces together, and the interference task would be articulatory suppression (repeating "the").*

In Article I, we conducted a systematic review of the verbal interference literature. Some of the studies included in the systematic review hypothesise a verbal rehearsal component akin to verbal working memory, as in task switching studies where the phonological loop is thought to be used to rehearse the relevant task and serve as self-cued reminders of task switches (Baddeley et al., 2001; Emerson & Miyake, 2003; Miyake et al., 2004). Other studies hypothesise verbal involvement that pertains more to the representational or structural properties of language, as in reorientation studies where inner speech supposedly serves as a common medium for integrating different types of information about the environment (Bek et al., 2013; Caffò et al., 2011; Hermer-Vazquez et al., 1999; Hupbach et al., 2007; Ratliff & Newcombe, 2008). We were interested in seeing what patterns we would find if we looked systematically at the method across fields, as dual-task interference effects have often been used in philosophical arguments about the role of language in thought (Carruthers, 2002; Clark, 2008). This article therefore presents a systematic review of the empirical literature that uses dual-task interference methods for investigating the on-line involvement of language in various cognitive tasks.

Our primary goals were:

- To provide a coherent overview to aid in understanding of what cognitive functions language may and may not be involved in.
- To provide methodological suggestions and recommendations for future studies in order to make results from different experiments more comparable.
- To provide theoretically motivated reasons for choosing one interference type over another.

2.1.2. Method

We assessed studies (N = 101) reporting at least one experiment with verbal interference and at least one control task (either primary or secondary). We excluded papers with an explicitly clinical, neurological, or developmental focus.

2.1.3. Results

The primary tasks identified were categorization (simple, N = 11, and complex, N = 5), memory (N = 15), mental arithmetic (N = 10), motor control (N = 2), reasoning (verbal materials, N = 8, and non-verbal materials, N = 12), task switching (N =16), theory of mind (N = 4), visual change (N = 6), and visuospatial integration and wayfinding (N = 12). We found four different kinds of interference tasks as well: syllable repetition, verbal shadowing, memory, and judgment.

2.1.3.1. Interference tasks

Syllable repetition, also known as articulatory suppression, was the most frequently used interference task in the review, found in 61 of the 101 studies. This kind of interference task is often paired with foot or finger tapping as the control interference task. The second most common interference task type was memory-based, found in 22 studies. Here, participants are asked to engage in covert rehearsal of verbal and non-verbal materials during the primary task with a subsequent memory test. This kind of interference task has the advantage over syllable repetition that it is easier to assess performance (through the memory test) but also the disadvantage that participants may be able to either encode the materials to be remembered non-verbally or in long-term memory, thus not interfering with the phonological loop. The two last interference tasks were verbal shadowing (13 studies) and verbal judgment (6 studies). In verbal shadowing, participants are asked to repeat a recording of fluent speech back as quickly as possible. This is often paired with clapped rhythm shadowing (same principle but with a recorded rhythm). It has the advantage of occupying both the "inner ear" and the articulatory organs but the disadvantage that performance is difficult to monitor and measure. In experiments using verbal judgment interference tasks, participants are asked to make phonological or semantic judgments of words (e.g., 'is this a real word?' or 'do these two words rhyme?'). On the one hand, judgment tasks are less "pure" in that they sometimes both target articulation (if participants give verbal responses), auditory imagery (as with rhyme judgments), and linguistic processing at a lexical level. On the other hand, interfering with semantic processing could be interesting if we hypothesise that a given primary task would rely on processing word meanings.

2.1.3.2. Primary tasks

Overall, the systematic review found that internal language is likely to play a facilitative role in memory and categorisation when items to be remembered or categorised have readily available labels, when inner speech can act as a form of behavioural self-cuing (inhibitory control, task set reminders, verbal strategy), and when inner speech is plausibly useful as "workspace", e.g., for mental arithmetic. There was less evidence for the role of internal language in cross-modal integration, reasoning relying on a high degree of visual detail or items low on nameability, and theory of mind. See Figure 3 for verbal interference effects by interference task and primary task. A common pattern seems to be that we observe specific effects of verbal interference when the primary task involves covert verbal rehearsal as a memory- or attention-guiding aid. With the possible exception of categorisation studies (Lupyan, 2009; Maddox et al., 2004; Minda et al., 2008), there is less evidence that verbal interference has a specific effect on tasks hypothesised to rely on language conceived as a structuring tool. This was for example the case with theory of mind studies where internal language is hypothesised to provide the necessary structure for thinking about false beliefs (Dungan & Saxe, 2012; Forgeot d'Arc & Ramus, 2011; Newton & de Villiers, 2007; Samuel et al., 2019).

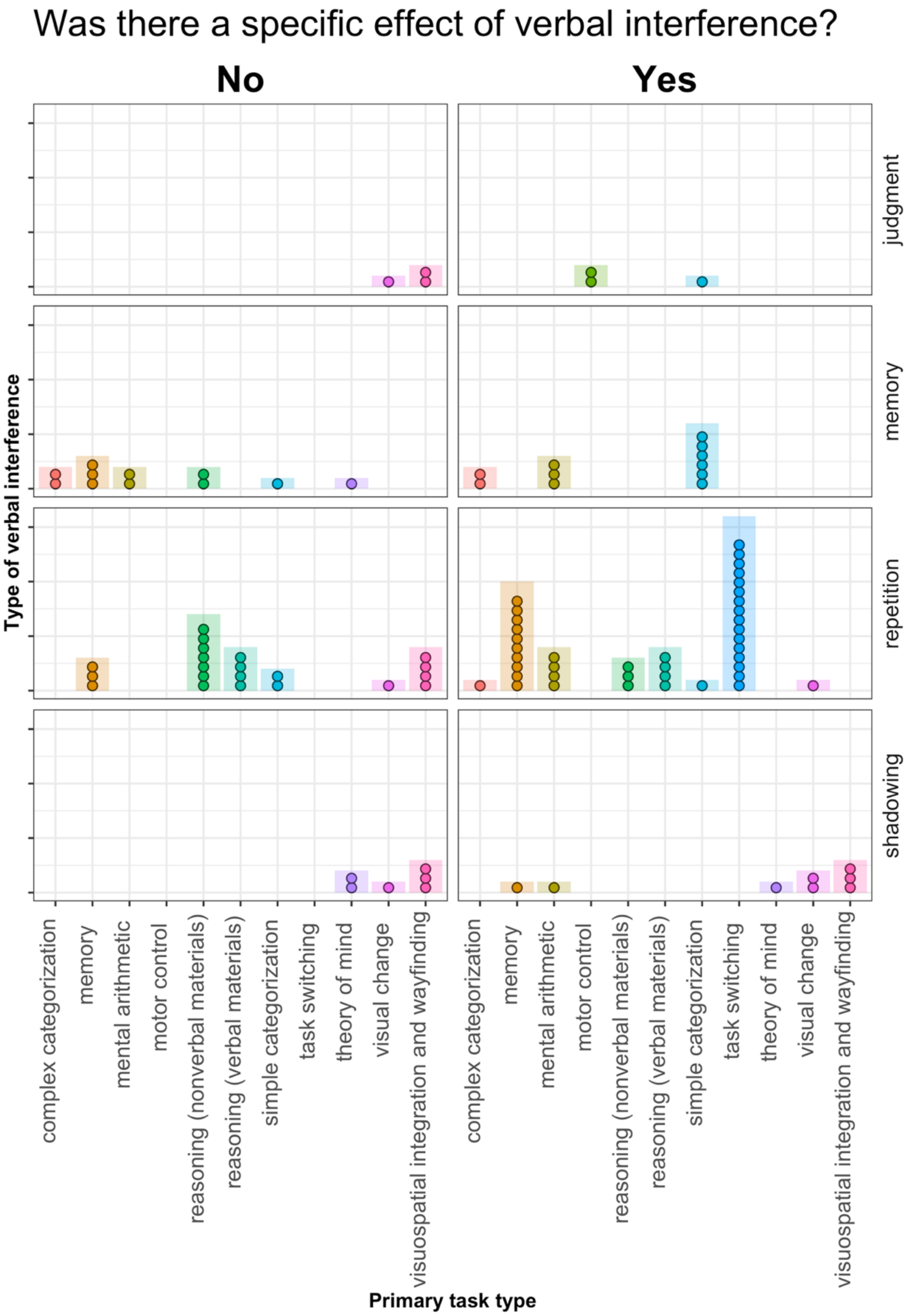


Figure 3. *Visualisation of the overall results where each point represents a study included in the systematic review. The 11 primary task categories are indicated on the x axis and by colour. Each row shows a different type of verbal interference. "Judgment" refers to judgment of verbal materials (for example rhyme), "memory" refers to the interference caused by a verbal memory task, "repetition" refers to repetition of simple syllables or words, and*

"shadowing" refers to the immediate repetition of continuously changing verbal material. Whether there was a specific effect of verbal interference (either compared with a non-verbal interference task or across different primary tasks) is indicated by the column-wise subplots in the plot grid.

2.1.4. Conclusions & implications

We found relatively robust evidence for verbal involvement in some areas and less robust (or even absent) evidence in other areas. There were both implicit and spontaneous language effects, such as on simple categorisation tasks, and effects of more explicit, language-based strategies, such as verbal self-cuing. We recommend that future studies should include both control primary tasks (where the theory predicts no verbal involvement) and control interference tasks. Without one or both of these, it is not possible to draw robust conclusions about covert language involvement. Interestingly, verbal interference sometimes had a *facilitative* effect on primary task performance, indicating that inner speech involvement could be associated with loss of visual detail, verbal overshadowing, and slower encoding (Brandimonte et al., 1992; Forgeot d'Arc & Ramus, 2011; Hitch et al., 1995; Pelizzon et al., 1999).

It is important to note that a lack of a specific verbal interference effect does not necessarily mean that language is not involved in that primary task – language could be involved in development or otherwise off-line. This seems for example to be the case with theory of mind where there is converging evidence that language plays an important role in development (Astington & Baird, 2005; Astington & Jenkins, 1999; Gagne & Coppola, 2017; Lohmann & Tomasello, 2003; Milligan et al., 2007; Pyers & Senghas, 2009; Slade & Ruffman, 2005) but little evidence for verbal interference in adults. To further refine the verbal interference method, which downregulates inner speech (i.e., makes it *less* likely that inner speech can be used), it could be fruitful to combine it with upregulating methods as is for example done with self-talk training in sport psychology interventions (i.e., making it *more* likely that specific inner speech strategies are used).

2.2. Article II: Cycling experiments

2.2.1. Background

Vygotskian accounts of inner speech hypothesise that it is used primarily for behavioural self-regulation (Diaz & Berk, 1992; Vygotsky, 1962). Inner speech helps focus attention on relevant aspects of the task at hand as well as the environment and has much the same effects as another person's instructions would. Self-regulation – both in terms of behaviour and motivation – plays an essential role in sports and physical endurance in general (Brick et al., 2016; Hyland-Monks et al., 2018; Kirschenbaum, 1987; McCormick et al., 2019). If inner speech is recruited to enhance physical endurance, then we would expect verbal interference to be detrimental to endurance cycling performance. To ensure that an effect of verbal interference is not just due to general dual-task attentional demands, it is necessary to include a control interference condition such as foot tapping or a visuospatial working memory task. This non-verbal interference condition should be equal in all other respects than the presence of *verbality* (preferably equally difficult and using comparable stimuli) (see Article I and Perry & Lupyan, 2013). This article presents two preregistered behavioural experiments[2] utilising the dual-task interference method to investigate the role of internal language in endurance control. In the experiments, participants were asked to do multiple one-minute sprints on an exercise bike while also performing a secondary interference task.

2.2.2. Hypotheses

[2] Experiment 1: https://osf.io/2ah7s; Experiment 2: https://osf.io/byfp3

- Cycling performance will decrease in both the verbal and non-verbal interference conditions compared to the control condition.
- If inner speech is required to maximise performance, we expect cycling performance to decrease significantly more in the verbal compared to the non-verbal interference condition.
- If there is no detectable dual-task effect on cycling performance, we expect to see a trade-off where there is instead a detrimental effect on the verbal or non-verbal simultaneous task.
- Participants who indicate high self-talk frequency and efficacy in the questionnaire will be more negatively affected by the verbal distraction task than other participants.

2.2.3. Method

In both experiments, we tested physically active participants' performance on an exercise bike (Experiment 1: N = 49; Experiment 2: N = 50) where they were asked to cycle 'as quickly as possible' for one minute while also having to remember either some verbal stimuli (hypothesised to interfere with inner speech) or some visuospatial stimuli (to control for attentional effects). We also included a control condition where participants did not have to remember anything while cycling. In both experiments, participants completed 12 cycling trials and 12 resting trials interleaved, and the order of the three conditions (verbal interference, visuospatial interference, and control) was randomised within four blocks. Participants were instructed to cycle as fast as they could and reach at least 70 % of their maximum heart rate during each cycling trial. A brief warm-up section served to illustrate the amount of effort required to reach the desired heart rate band.

The interference tasks differed between the two experiments: in Experiment 1, participants were asked to remember either six letters and numbers (verbal interference) or the locations of six letters and numbers on a grid (visuospatial interference). In both conditions, the stimulus presentation was the same – six letters and numbers appeared one by one for one second each in different locations on a 6x6 grid (see Figure 4).

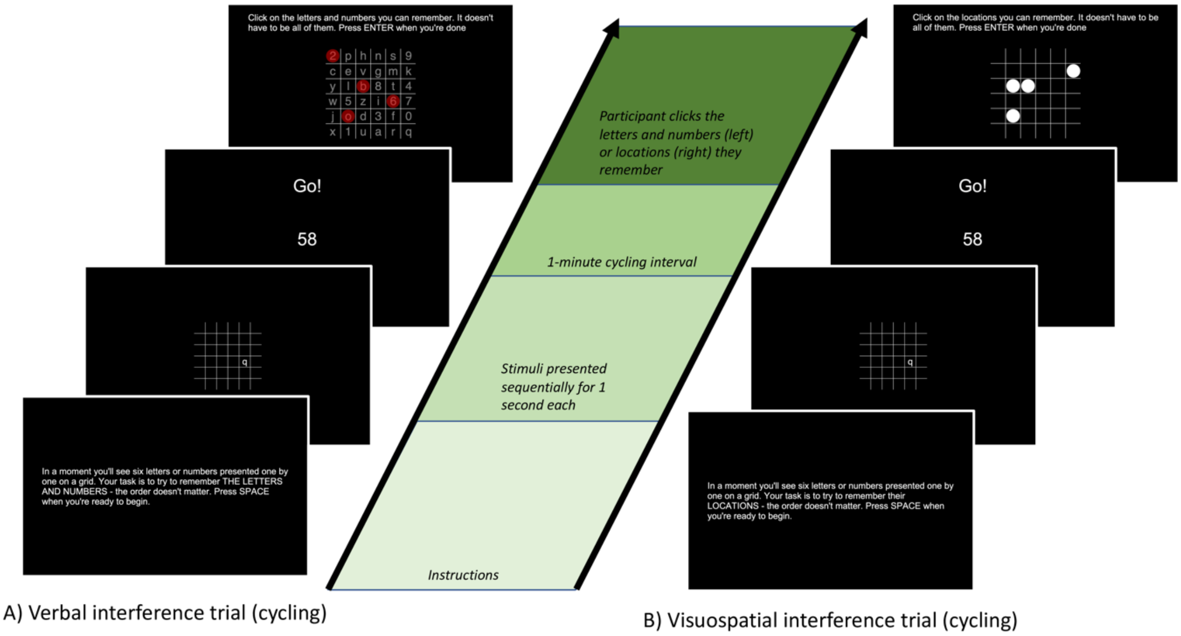


Figure 4. *Sketch of the Experiment 1 procedure on a cycling trial with verbal interference trial (A) and a cycling trial with visuospatial interference (B). In the experiment, six letters and numbers were presented sequentially for one second each in random locations on the 6x6 grid, similar to the way the letter 'q' is presented on the figure.*

During the resting trials, participants also sometimes had to remember letters and numbers or locations on the 6x6 grid. There were some issues with the two interference tasks used in Experiment 1, however. First, the visuospatial interference task was much more difficult than the verbal interference task (visuospatial condition median accuracy = 3/6; verbal condition median accuracy = 6/6). Second, given the relatively long delay between encoding and reproducing the stimuli (one minute), participants did not necessarily rely on working memory if they were able to transfer the information to long-term memory. This means that the phonological loop might not have been continuously occupied. For these reasons, and to test the robustness of the results, we conducted a second experiment with different interference tasks. In Experiment 2, the verbal interference task was continuous 2-back matching of auditorily presented nonsense words (designed to minimise the possibility of visual encoding, e.g., picturing a pen if the word was "pen"), and the visuospatial interference task was continuous 2-back matching of coloured polygon shapes (designed to minimise the possibility of verbal encoding, e.g., naming the shapes "triangle" or "blue one"). See Figure 5. Both experiments were custom-written in PsychoPy software (Peirce et al., 2019).

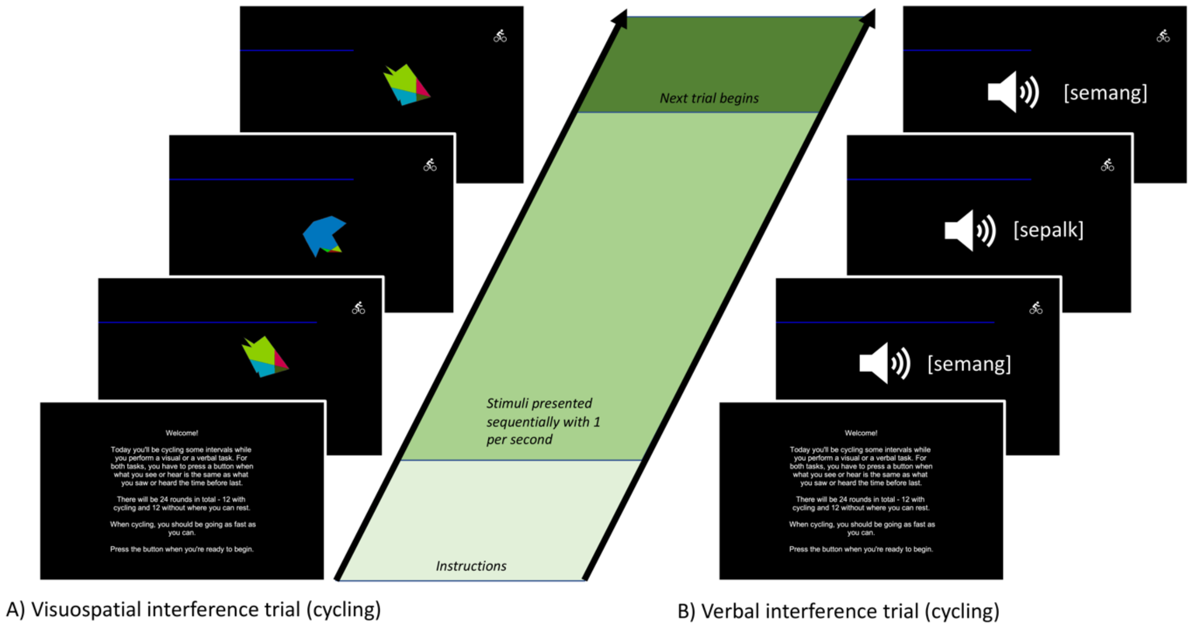


Figure 5. *Sketch of the Experiment 2 procedure on a cycling trial with visuospatial interference (A) and a cycling trial with verbal interference (B). Polygon positions varied, but each individual polygon was always displayed in the same location on the screen. The nonsense words are presented orthographically on the figure for the reader's benefit – they were solely presented auditorily to the participants during the experiment.*

2.2.4. Analysis & results

We found that participants performed worse (i.e., cycled a shorter distance) in the verbal interference condition compared with the control condition in both experiments (Exp 1: d = 0.29; Exp 2: d = 1). Participants were numerically slower during verbal interference than during non-verbal interference in both experiments. However, there was only a significant difference between the verbal and the visuospatial interference tasks in the second experiment (Exp 1: d = 0.22; Exp 2: d = 0.43), probably due to the nature of the interference tasks. See Figure 6 for participants' cycling performance (meters cycled) over the duration of the experiment in the three conditions in Experiment 1 and Figure 7 for the same in Experiment 2. Note that Figure 7 displays performance as cadence instead of meters cycled because we used a different way to measure cycling output.

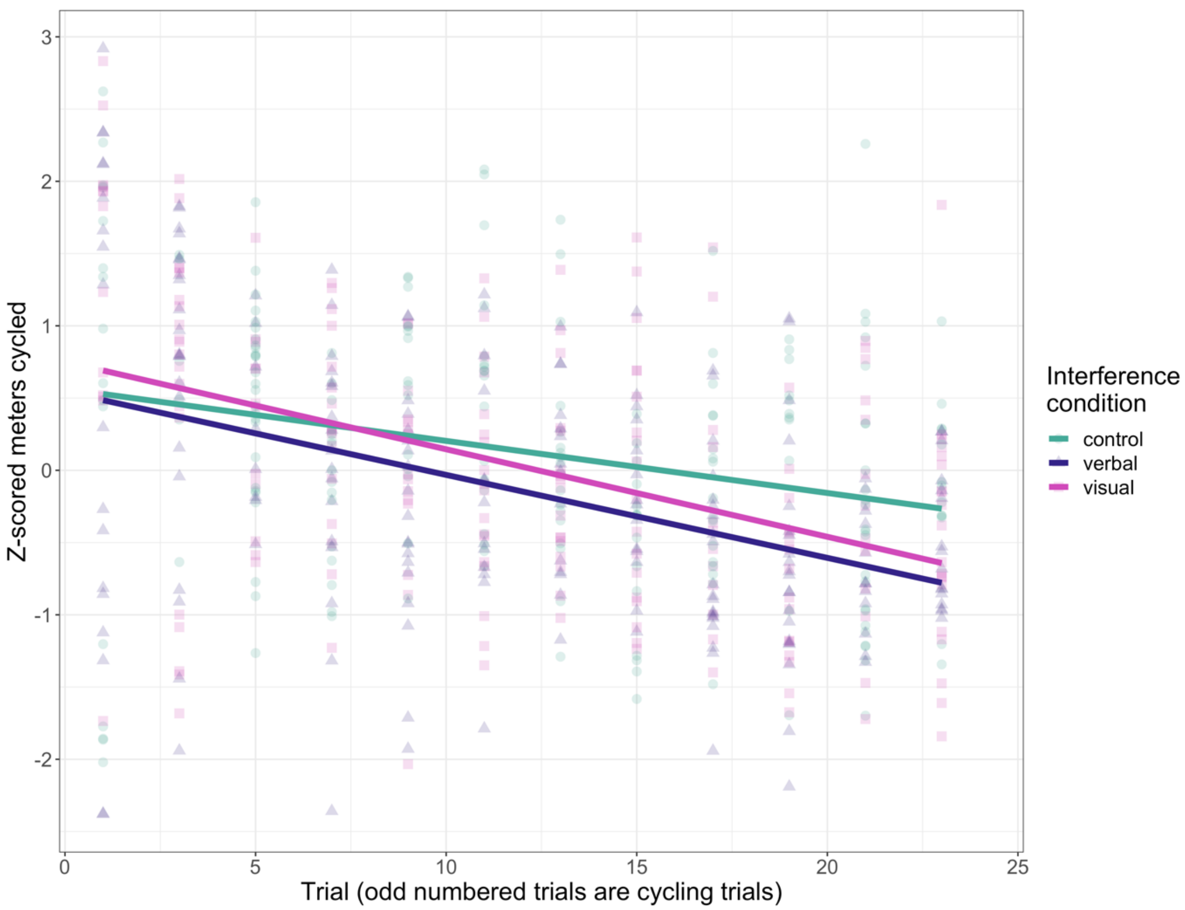


Figure 6. *Z-scored meters cycled across all trials in Experiment 1. Green line and dots represent control trials, blue line and triangles indicate verbal interference trials, and pink line and squares represent visuospatial interference trials.*

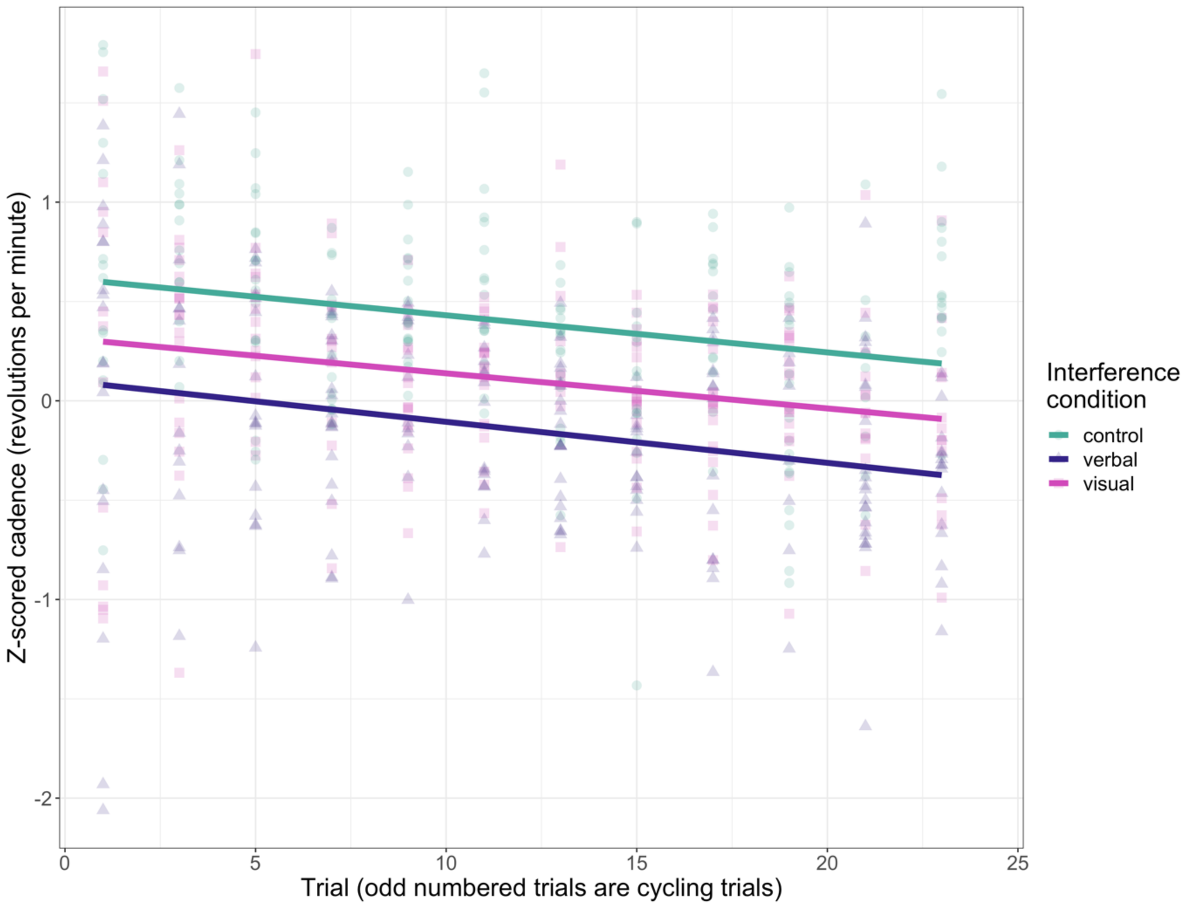


Figure 7. *Z-scored cadence across all trials in Experiment 2. Green line and dots represent control trials, blue line and triangles indicate verbal interference trials, and pink line and squares represent visuospatial interference trials.*

As one of our preregistered hypotheses concerned the effects of experienced self-talk efficacy ('What effect does self-talk usually have on your performance?') and self-talk frequency ('How often to you talk to yourself while exercising?'), we tested whether the degree of verbal interference varied with either of these factors. There were no significant relationships in either Experiment 1 or Experiment 2.

2.2.5. Conclusions & implications

The findings support the idea that inner speech is involved in behavioural control and especially control of physical endurance. The specific mechanisms involved could be behavioural self-cuing, inhibitive control, and motivation. It appears that under verbal interference, participants were less able to use their inner voice to focus their attention on the task demands and inhibit their propensity to slow down, and this had detrimental effects on their cycling performance. While the effects of self-talk have been extensively investigated in sport psychology, this was the first study to test these effects directly using the interference method. This study also contributes an important comparison between different types of verbal interference and indicates that memory-based interference is less disruptive than 2-back matching. However, dual-task interference is inherently a coarse method that leaves us limited in how much we can say about what it is about inner speech that helps. A combination of this dual-task interference method and the self-talk interventions often used in sport psychology research would be a promising avenue for future research as it would allow us to up- and down-regulate different ways of talking to oneself. This could also help disentangle the (lack of) correspondence between verbal interference effects and self-reported inner speech efficacy in individuals. To get a more reliable measure of self-talk frequency, future studies could use DES in a sport setting (similar to Dickens et al., 2018).

2.3. Article III: Attention regulation

2.3.1. Background

As previously discussed, some of the most frequently self-reported functions of inner speech are self-regulation and motivation (Alderson-Day et al., 2018; Alderson-Day & Fernyhough, 2015b; Morin et al., 2011; Uttl et al., 2011). As we saw in Article II, there is evidence that people do benefit from being able to talk to themselves to control physical endurance. This benefit may stem from the role of inner speech in impulse control, which is relevant in many areas other than physical endurance (Dunbar & Sussman, 1995; Tullett & Inzlicht, 2010; Wallace et al., 2017). There is indeed evidence that people generally talk to themselves to stay focused on a task that is tedious or to refrain from making inappropriate responses (Tullett & Inzlicht, 2010). Inner speech appears to be especially recruited under challenging circumstances, when learning new skills or when a high degree of top-down control is necessary (Emerson & Miyake, 2003; Kray et al., 2008). In Article III, we explored a paradigm which was intentionally so boring that participants would presumably need a lot of top-down control – and thereby potentially inner speech – to stay focused. To measure the degree of control exerted, we also asked participants to report how their minds wandered away from the task.

In situations demanding top-down control of attention, mind-wandering is associated with failures to monitor task performance, thus leading to more errors (Smallwood et al., 2007). The literature on mind-wandering has generally not been concerned with the specific modality in which inner experience takes place, but rather whether it is task-relevant or not. To get this kind of information, mind-wandering research often uses an experience sampling method that is less resource-intensive – but also less sophisticated – than Descriptive Experience Sampling (Hulburt and colleagues). It represents a compromise between data quantity and data quality. Descriptive Experience Sampling is not possible with a very large number of participants because each participant has to be interviewed iteratively by at least two experimenters (Hurlburt & Heavey, 2018). Experience sampling during a specific experiment with set questions and options probably yields more reliable data than an ordinary questionnaire but less reliable than Descriptive Experience Sampling (Alderson-Day & Fernyhough, 2015a; Hurlburt & Heavey, 2015). In Article III, we used questionnaire-based experience sampling and compared our results to those obtained by Descriptive Experience Sampling and by the Varieties of Inner Speech Questionnaire-Revised (VISQ-R; Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011).

We were interested in whether people need to talk to themselves to stay focused on a task that requires nothing but their attention. To investigate this, we tested if the presence or absence of task-relevant thought in general and task-relevant inner speech in particular was predictive of response time to infrequently occurring prompts. We conducted a preregistered online experiment and a preregistered replication because we deviated substantially from the original registered analysis plan. Data, preregistrations, and analysis and experiment code are available on the Open Science Framework[3].

2.3.2. Hypotheses

Our preregistered hypotheses were as follows:

- Task-relevant inner experience will generally be associated with faster reaction times to the prompt.
- Specifically, task-relevant inner speech will be associated with faster reaction times than other types of inner experience.
- The proportions of types of inner experience will resemble those found in other experience sampling studies.
- Exploratory: Self-regulatory inner speech may be more important as the experiment progresses, due to build-up of boredom/fatigue. If this were the case, we would predict an interaction between the inner speech factor and time with the difference between task-relevant inner speech and task-relevant other experience becoming more pronounced over time.
- Exploratory: Response time variance will be lower for task-relevant inner speech trials.

[3] https://osf.io/jgx7m/

2.3.3. Method

In both the original experiment and the replication, we measured response times to an infrequently occurring stimulus (a black dot appearing at 1-3 minute intervals) and subsequently asked participants to report on the character of their inner experience at the time the stimulus appeared. See Figure 8 for a sketch of the experiment procedure common to both experiments.

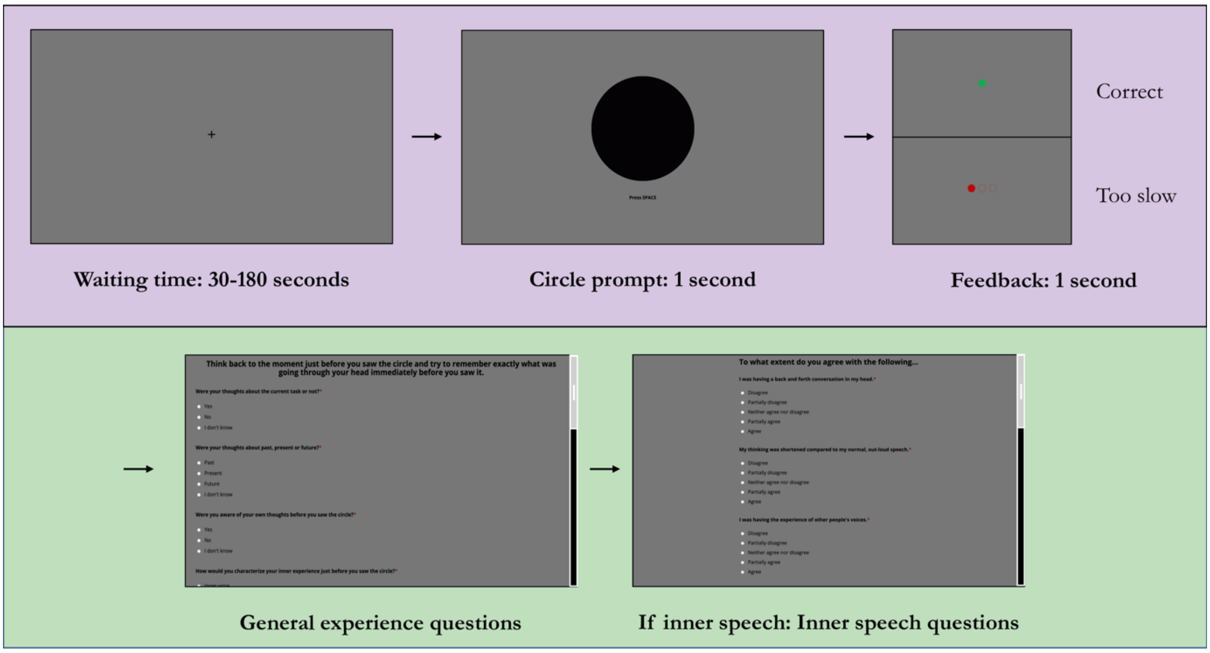


Figure 8. *Illustration of the experiment progression.*

After responding to the circle prompt, participants were asked some questions related to their inner experience at the time of the prompt (see Table 1). The five categories of inner experience they were asked about were inspired by the experience types found in Descriptive Experience Sampling. If participants answered that their inner experience was best characterised as "Inner voice", they were asked some follow-up questions inspired by the VISQ-R (how they experienced their inner voice as dialogic, condensed, and evaluative, as well as whether they had the experience of other people's voices).

Table 1. *The questions posed to participants after each circle prompt.*

Question (inner experience)	**Options**
Were your thoughts about the current task or not?	"Yes", "No", "I don't know"
Were your thoughts about past, present or future?	"Past", "Present", "Future", "I don't know"
Were you aware of your own thoughts before you saw the circle?	"Yes", "No", "I don't know"
How would you characterise your inner experience just before you saw the circle?	"Inner voice", "Inner seeing", "Unsymbolised thinking", "Sensory awareness", "Feelings"
Question (inner speech-specific)	**Options**
I was having a back and forth conversation in my head.	"Disagree", "Partially disagree", "Neither agree nor disagree", "Partially agree", "Agree"
My thinking was shortened compared to my normal, out-loud speech.	*Same as the above.*
I was having the experience of other people's voices.	*Same as the above.*
I was evaluating my behaviour using my inner speech.	*Same as the above.*

When the participant had responded to all circle prompts (eight in the original experiment, 12 in the replication experiment), they were also asked whether they talked to themselves to stay focused throughout the experiment. Both experiments were conducted via the participant-recruitment platform Prolific using custom-written software and the JavaScript library jsPsych

(De Leeuw, 2015). We analysed data from 212 participants in Experiment 1 and 222 participants in the replication.

2.3.4. Analysis & results

Participants most frequently reported that their experience had been in a verbal format in both the original and the replication experiment. See Figure 9.

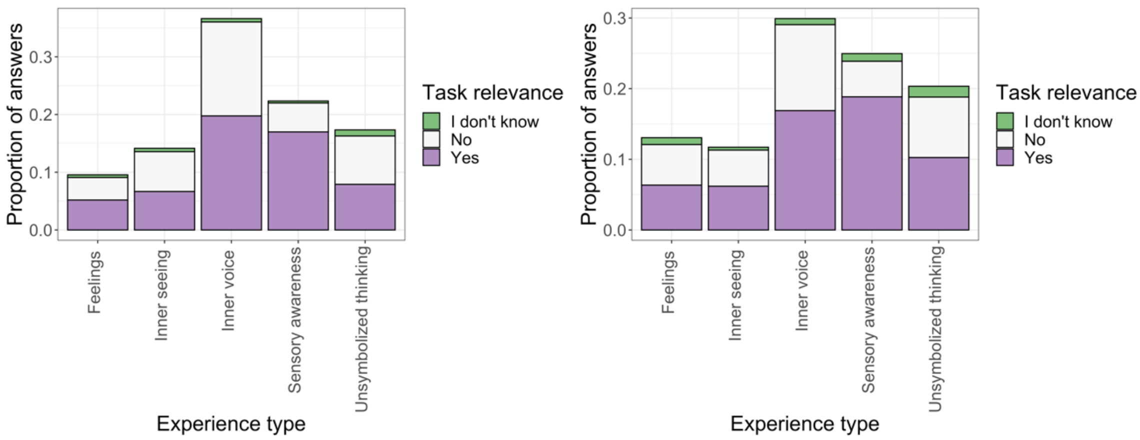


Figure 9. *Experience types in response to prompts in the original experiment (left) and in the replication (right).*

For the more in-depth inner speech questions, the distributions of answers were comparable to those found in other VISQ-R studies (Alderson-Day et al., 2018). See Figure 10.

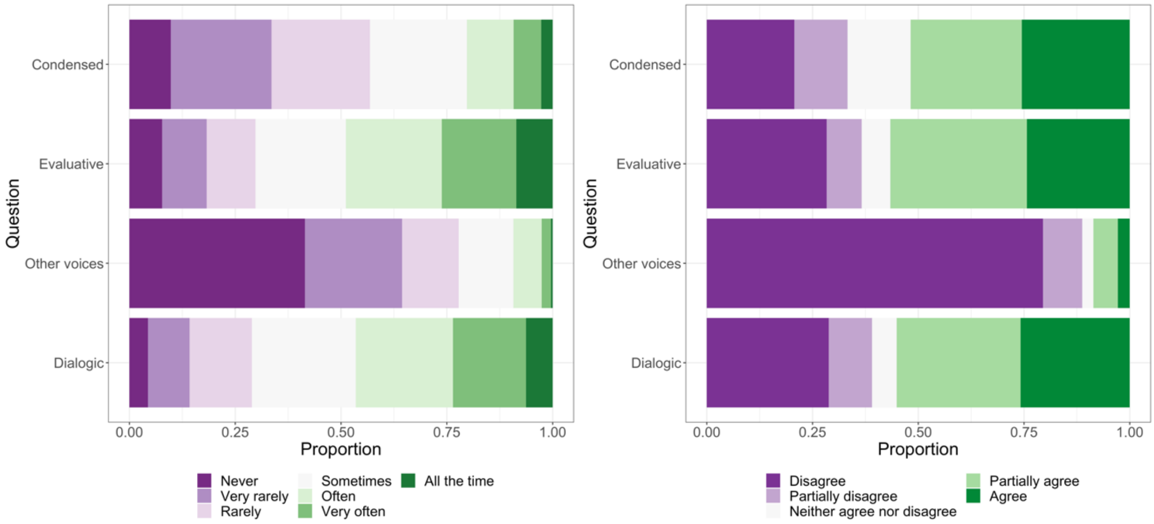


Figure 10. *Distributions of answers to the inner speech questions in Alderson-Day et al. (2018) on the left and our original study on the right (our results did not differ substantially between the original experiment and the replication). Note the different Likert scales.*

With generalised linear mixed-effects models fitted to a Gamma distribution, we found significant effects of task relevance but no interaction with inner speech. However, using a hierarchical Bayesian analysis method (not preregistered), we found that trials preceded by task-relevant inner speech additionally displayed lower standard deviation and lower mode independently of the main effect of task relevance. Because we had deviated from the planned analyses in Experiment 1, we conducted a second experiment to replicate our results and test their robustness. For Experiment 2, we only used the hierarchical Bayesian analysis method and found both a main effect of task relevance and an interaction effect between inner speech and task relevance. See Figure 11 for an illustrating of the crucial interaction effect. Neither the effect of task-relevant experience nor of inner speech became more important as the experiments progressed.

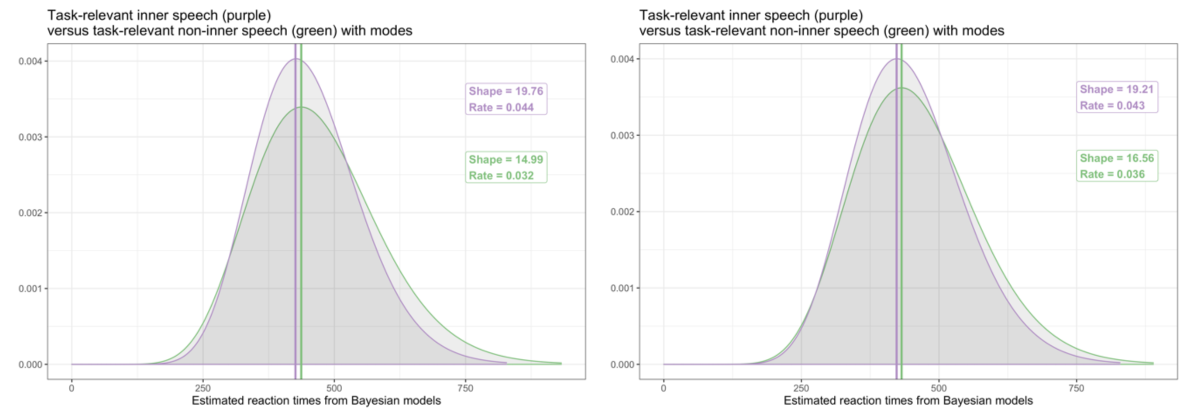


Figure 11. *The interaction effect (task-relevant inner speech trials (purple) versus task-relevant non-inner speech (green)) in the original experiment (left) and in the replication (right). The density plots represent posterior estimates from the Bayesian analyses, and the shape and rate parameters define the Gamma distributions. Response times preceded by task-relevant inner speech displayed both lower standard deviation and lower mode.*

2.3.5. Conclusions & implications

Our aim with this study was to explore 1) how inner speech is involved in sustained attention, and 2) the usefulness of a compromise between questionnaire-based measures of inner speech and Descriptive Experience Sampling. Our results add support to the hypothesis that inner speech serves a functional role in top-down attentional control, as trials preceded by task-relevant inner speech were associated with faster and less distributed response times. The response time differences are small, but it is nevertheless interesting that inner speech might have effects on such simple mechanisms as responding to a prompt with a button click. The interaction between inner speech and task relevance is particularly important as it indicates that talking to yourself is not *in itself* helpful. It needs to be about the task. Conversely, task-relevant inner experience appears to be in itself helpful, but the effect can be enhanced if task-relevant experience is accompanied by inner speech.

We were relatively successful in our pursuit of a compromise between questionnaire-based methods and Descriptive Experience Sampling. Our participants reported inner speech more often than participants generally do in DES, which may have been due to the lack of iterative training in reporting experience and the fact that participants could only choose one experience type in our experiments. Participants in DES often misidentify episodes of inner speech before they have had iterative training, and experience types are not mutually exclusive in DES (Hurlburt & Akhter, 2006). We run the risk of participants reporting their preconceived notions of the characteristics of their experience rather than pristine inner experience. This risk is perhaps increased by us not giving careful instructions and explanations for what the different items in the sampled-moment questionnaire meant. However, it could also be the case that the experiment circumstances meant that inner speech episodes were more likely than they are in normal life where external stimuli and social interactions often demand attention. The features of inner speech that we asked about which were inspired by the VISQ-R showed very similar distributions across our two studies and compared with previous studies (Alderson-Day et al., 2018), suggesting methodological robustness and comparability with other studies.

2.4. Article IV: Anendophasia

2.4.1. Background

It is very often assumed both by individuals and by researchers that inner speech is a ubiquitous phenomenon of inner life. However, recent debates taking place on social internet sites such as Twitter and Reddit have provided some nuance. Many people participating in these debates claim that they never experience inner speech, at least not in any identifiable natural language, and that they instead think in images or in “ideas”. These people’s experiences to some extent mirror those of people with aphantasia, the inability to engage in visual mental imagery. Indeed, self-reported visual imagery and verbal imagery correlate both in terms of participants’ ratings of vividness of imagined stimuli (Dawes et al., 2020) and reported tendency/propensity to engage in visual and verbal imagery generally (Roebuck & Lupyan, 2020)

Unlike aphantasia, which is now relatively well-described (Dance et al., 2022; Jacobs et al., 2018; Keogh et al., 2021; Keogh & Pearson, 2018), the lack of inner speech has not as yet received much research attention. There are important questions to be answered: could these people be mistaken about their inner experience? Do they just have a different understanding of what "inner speech" means? Is the feeling of a lack of inner speech categorical or a continuum? If they are right that they do not have inner speech, does this have any consequences for how they behave and solve problems? In this paper, we set out to investigate these questions.

2.4.2. Method

We recruited participants with very high and very low verbal representation scores on the Internal Representations Questionnaire (Roebuck & Lupyan, 2020) which asks participants about the characteristics of their inner experience. For example, an item with a high loading on the verbal representation factor could be 'I think about problems in my mind in the form of a conversation with myself' while an item with a high loading on the visual representation factor could be 'I often enjoy the use of mental pictures to reminisce'. After excluding 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 relatively high verbal representation scores on the IRQ (top 40%-ile) and 46 participants with relatively low verbal representation scores (bottom 16%-ile). The two groups did not differ in terms of age, gender, dyslexia, education level, or native language. In counterbalanced order, the 93 participants completed a verbal working memory task, a pictorial rhyme judgment task, an arithmetic task that included cued and un-cued switching between two different rules (adding 3 or subtracting 3), and a visual discrimination task. Participants also completed a custom questionnaire about their experience with inner speech or lack thereof, auditory imagery, and what they believe other people's experience is like. A full set of questions can be seen in the accompanying manuscript. The behavioural experiments were run through the participant-recruitment platform MTurk using custom-written software and the JavaScript library jsPsych (De Leeuw, 2015).

2.4.2.1. Verbal working memory

Participants were asked to remember and reproduce five words in sequential order. These five words were either phonologically similar but not orthographically similar ("bought", "sort", "taut", "caught", and "wart"), orthographically similar but not phonologically similar ("rough", "cough", "through", "dough", "bough"), or from a control set ("plea", "friend", "sleigh", "row", "board"). See Figure 12 for a sketch of this experimental task.

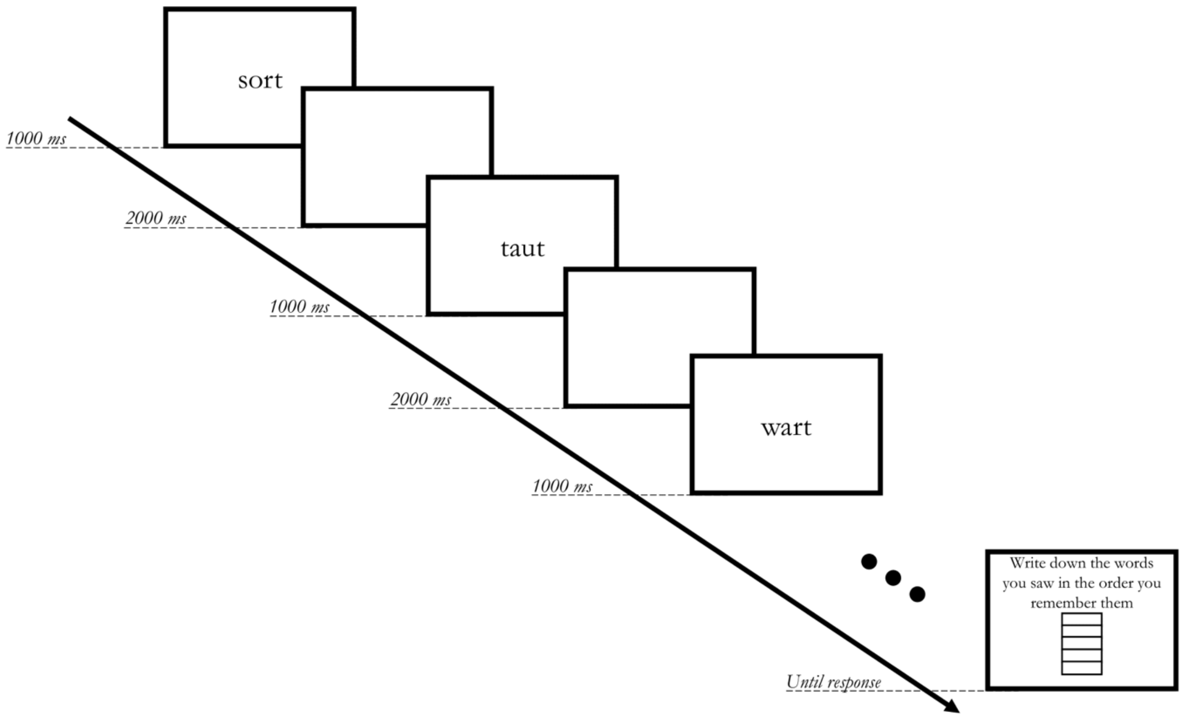


Figure 12. *A sketch of the procedure on a phonological similarity trial in the verbal working memory experiment. Participants always saw five words – three from the phonological similarity set are shown here for clarity.*

2.4.2.2. Rhyme judgments

Participants saw two images on each trial and were asked to judge whether the word for the two images rhymes or not (e.g., images of a house and a mouse). The rhyme pairs included both orthographic rhymes ("cat" and "hat") and non-orthographic rhymes ("drawer" and "door"). See Figure 13 for a sketch of the rhyme task.

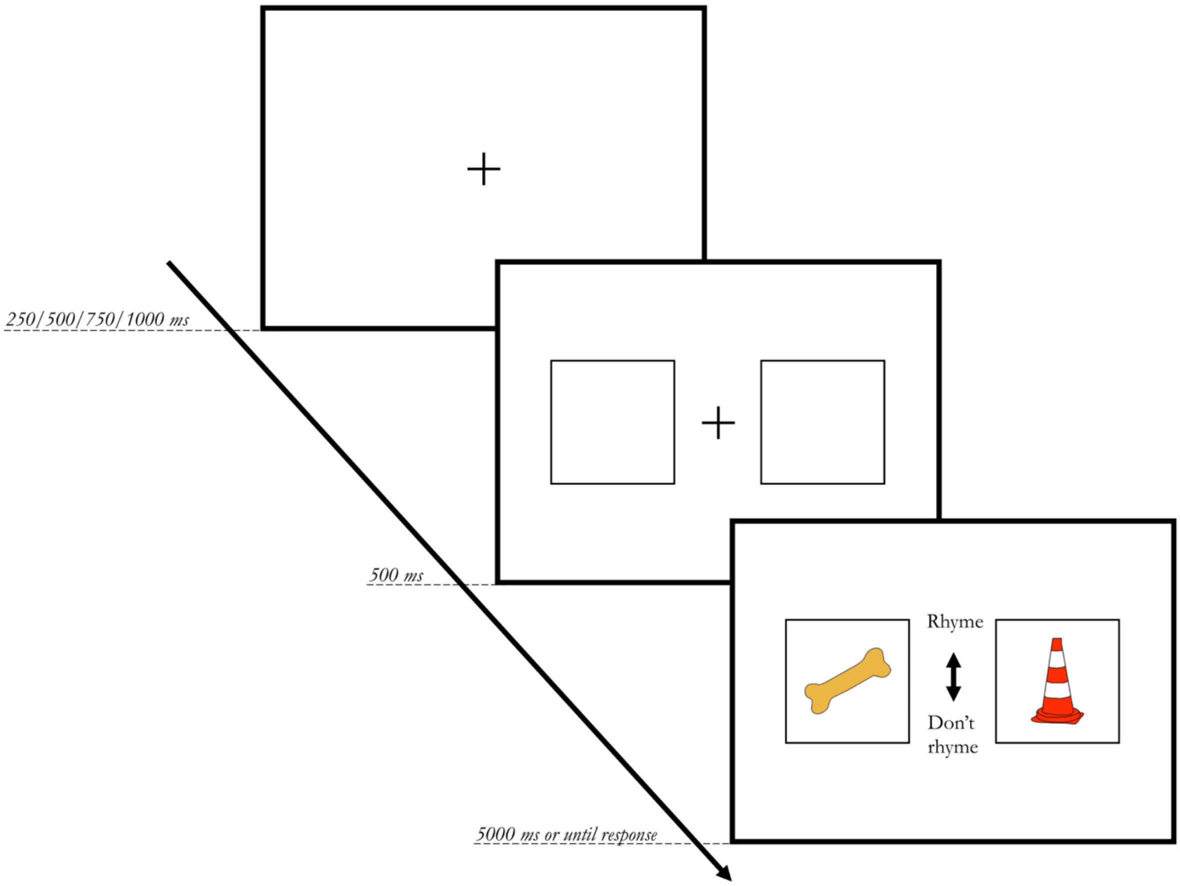


Figure 13. *A sketch of the procedure on an orthographic rhyme trial in the rhyming experiment.*

2.4.2.3. Task switching

Participants were asked to complete a series of arithmetic problems in five different conditions: blocked addition, blocked subtraction, colour-cued (red or blue) switching between subtraction and addition, symbol-cued (+ or -) switching between subtraction and addition, or un-cued switching between subtraction and addition. See Figure 14 for a sketch of the three switching conditions.

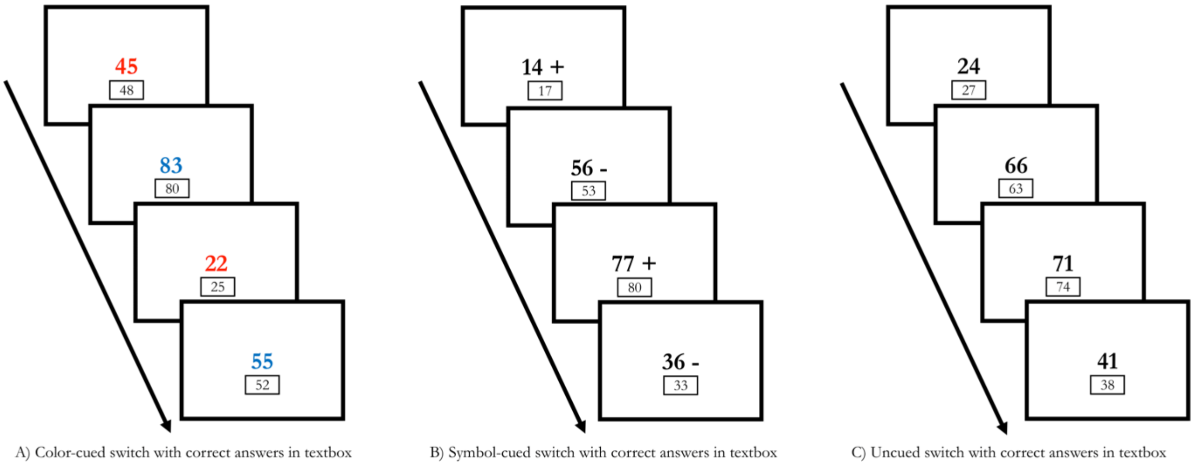


Figure 14. *Examples of four trials in each of the three switching conditions in the task switching experiment. Figure 14A shows the colour-cued switch condition, Figure 14B shows the symbol-cued switch condition, and Figure 14C shows the un-cued switch condition. For all three, the illustration includes correct answers in the textbox.*

2.4.2.4. Category judgments

Participants saw two black silhouette images of cats and dogs on each trial and were asked to judge either whether the animals were the same category (category condition) or physically identical (identity condition). See Figure 15 for a sketch of the categorical and identity judgment trials.

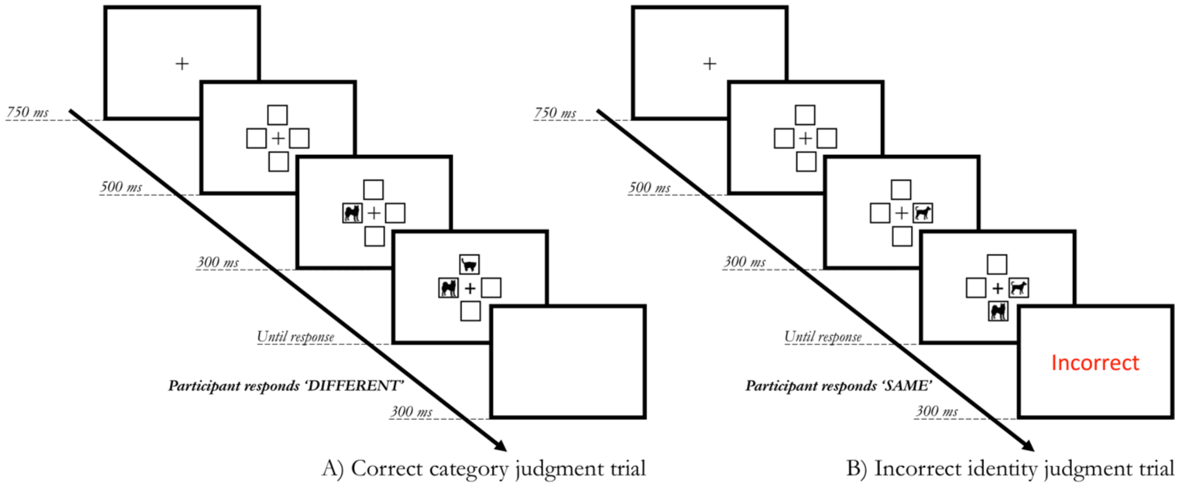


Figure 15. *A sketch of the procedure in the same/different judgment experiment. Figure 15A shows a correct category judgment trial, and Figure 15B shows an incorrect identity judgment trial.*

2.4.3. Analysis & results

2.4.3.1. Verbal working memory

Participants with more inner speech were generally better at remembering series of words. Contrary to our predictions, the phonological similarity effect (greater difficulty remembering phonologically similar words compared to control words) was not larger for the group reporting high levels of inner speech. When asked about their strategies for remembering the words, many participants with less inner speech reported that they had just remembered the starting letters of the words instead of the full words – a strategy that would be most efficient for the orthographic similarity set where the difference between the two groups was indeed reduced. This first-letter strategy was much more common for the group with less inner speech than for the group with more inner speech. In the latter group, the most common strategy was creating a story or a sentence with the words in order (relying on verbal or visual imagery, or both). Effects of different strategies have been shown in previous working memory studies as well, though not connected to inner speech propensity (Logie et al., 1996). See Figure 16.

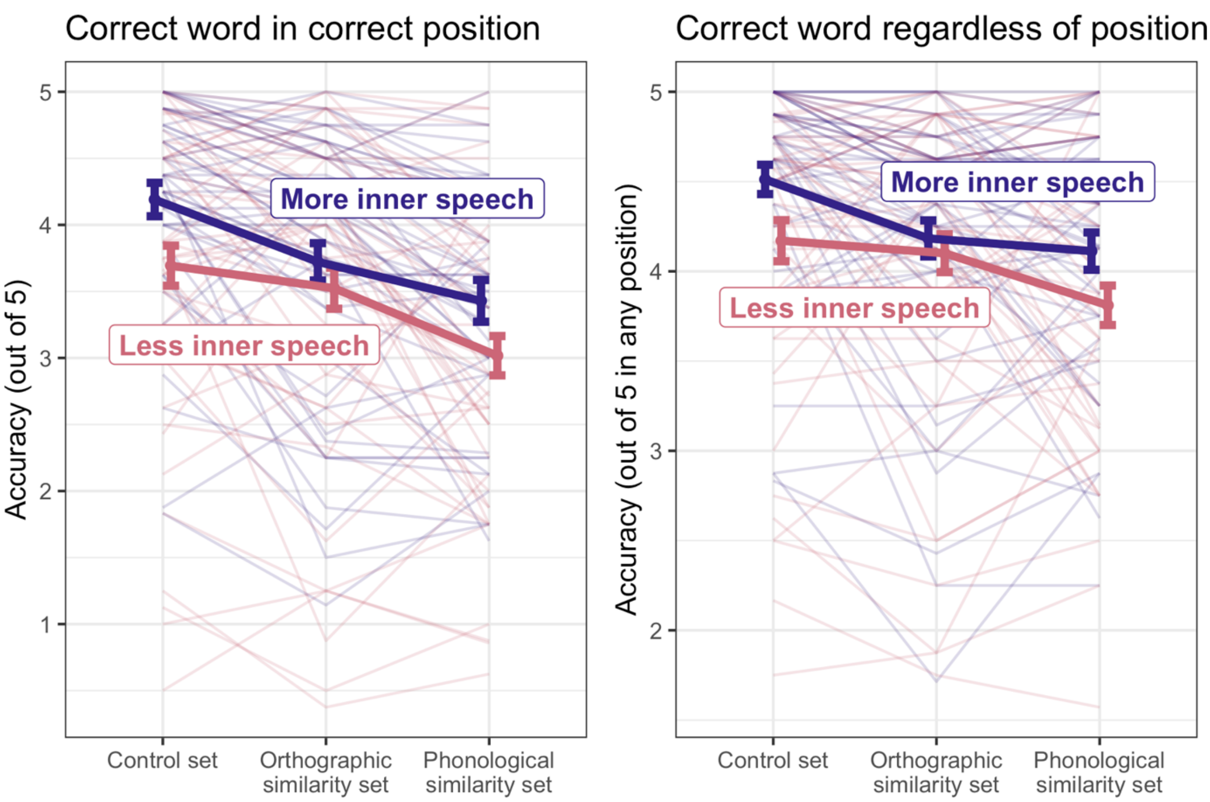


Figure 16. *Verbal working memory performance across the three word sets for the two groups of participants (reporting more and less inner speech). The figure on the left shows score when only correct items in correct position counted, and the figure on the right shows score when correct items regardless of position were counted. Error bars indicate 95% confidence intervals.*

Interestingly, the differences between the two participant groups were greatly diminished for the participants who reported talking out loud to help them remember the words (see Figure 17). The proportions of participants who reported talking out loud did not differ significantly between the two groups.

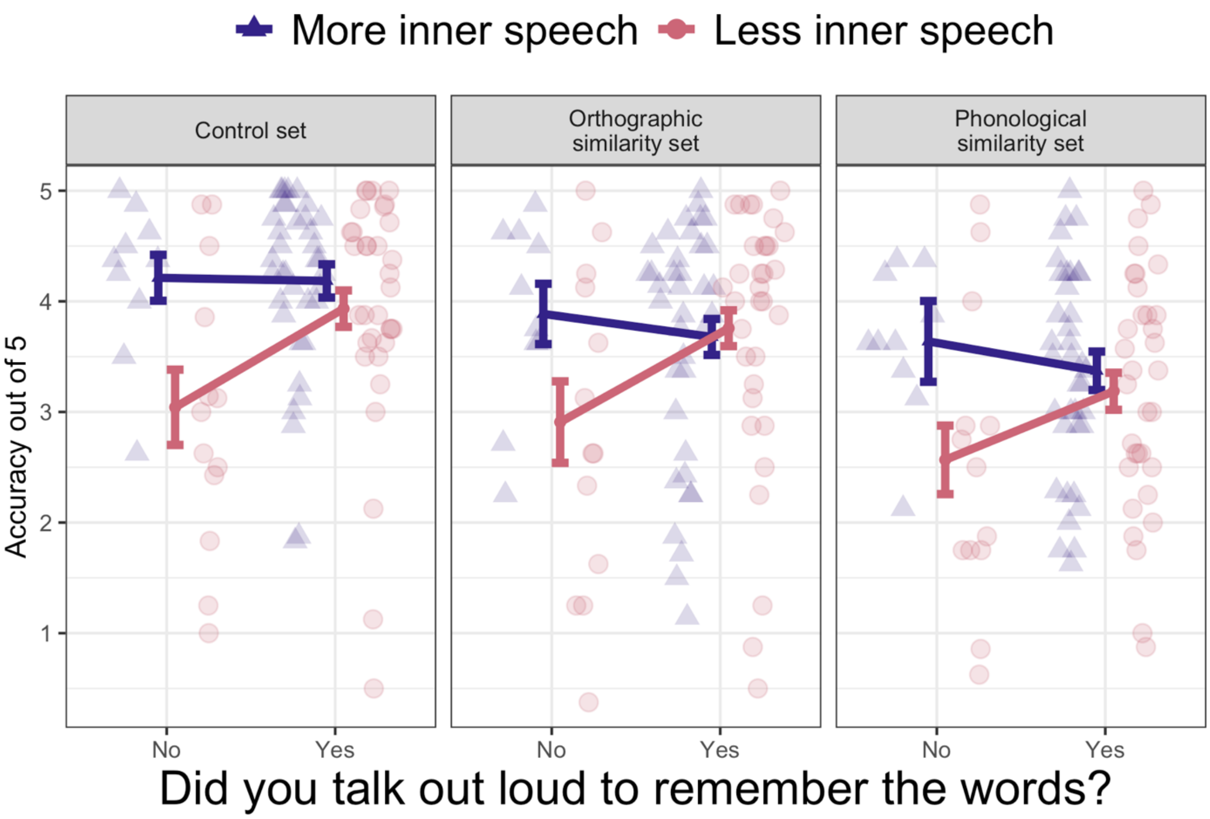


Figure 17. *Memory performance (correct item and correct position) as a function of word set, participant group, and whether participants reported talking out loud to help them remember the words. Error bars indicate 95% confidence intervals.*

2.4.3.2. Rhyme judgments

Participants with more inner speech were generally better but not faster at making rhyme judgments. There were no interactions involving orthographic rhyme or non-orthographic rhyme. See Figure 18.

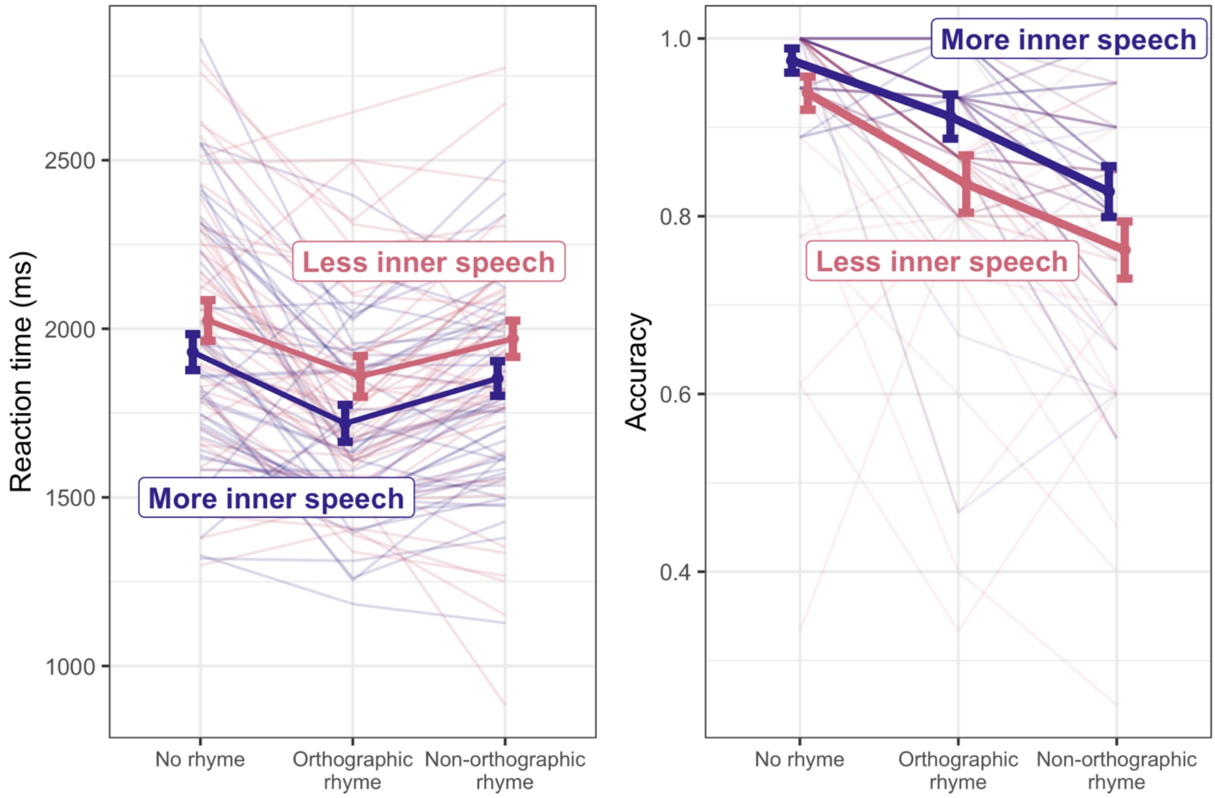


Figure 18. *Reaction time and accuracy in the rhyme judgment experiment as a function of rhyme type and participant group. Error bars indicate 95% confidence intervals.*

Interestingly, we saw similar effects as in the verbal working memory experiment of participants reporting talking out loud to help them solve the task. In the rhyme judgment experiment, participants with less inner speech who talked out loud were better at making both non-orthographical rhyme judgments and orthographical rhyme judgments (reaching the same level as participants with more inner speech). See Figure 19. Once again, the proportions of participants who reported talking out loud did not differ significantly between the two groups.

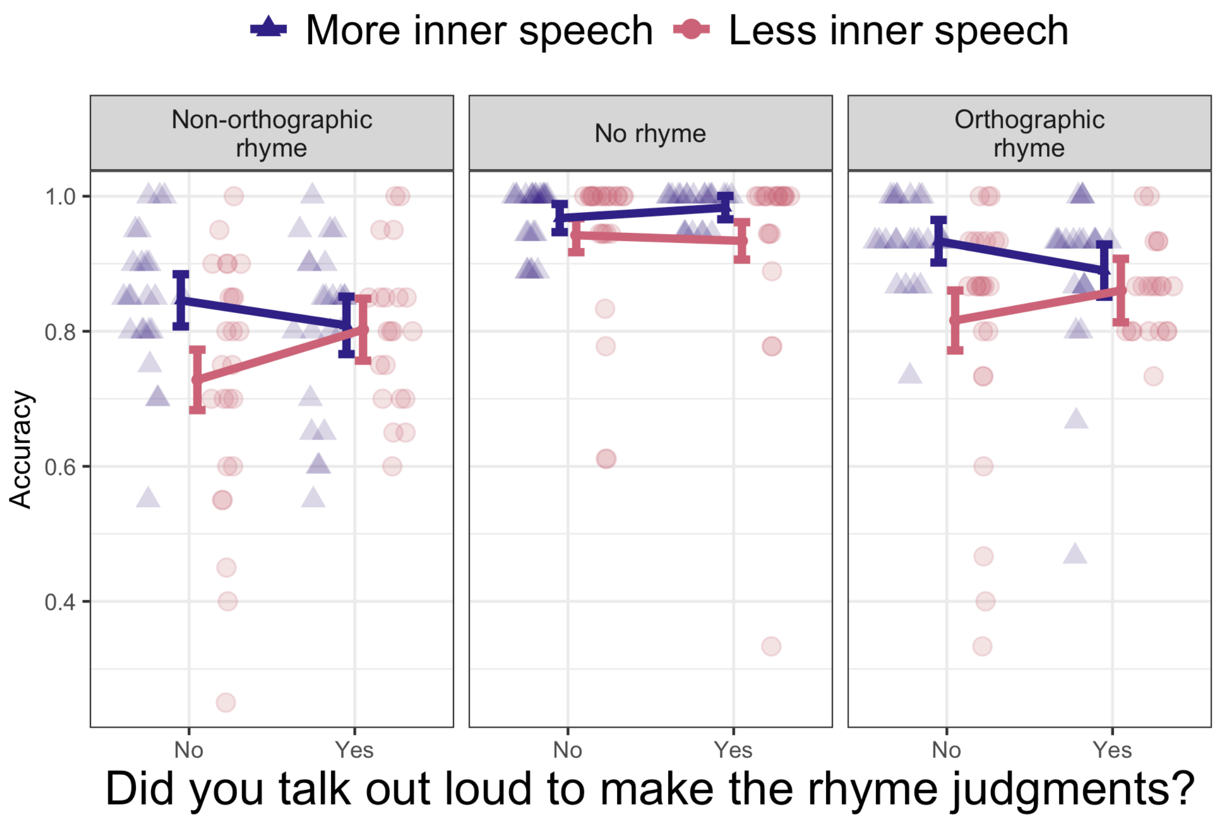


Figure 19. *Accuracy on rhyme judgments as a function of rhyme type, participant group, and whether participants reported naming the pictures out loud to make the rhyme judgments. Error bars indicate 95% confidence intervals.*

2.4.3.3. Task switching

There were no differences between the two groups on this experiment for either reaction time or accuracy although we did confirm the general costs of switching in terms of both accuracy and reaction time. There were also no notable differences between participants who reported talking out loud to remember the correct operation and participants who did not.

2.4.3.4. Same/different judgments

On identity judgment trials, when the correct answer was “different” (e.g., dog_1 and dog_2), we expected that participants with more inner speech would experience more category interference and thereby respond slower than participants with less inner speech would. We did not find such a difference, and there were also no general differences between participants with more inner speech and less inner speech across all trials. However, we did find that the participants taken as one group displayed the category interference effect on within-category identity judgment trials, thus replicating previous findings (Lupyan et al., 2010). There were no notable differences between participants who reported talking out loud during the same/different experiment and participants who did not.

2.4.3.5. Questionnaire

There were several striking differences in how participants with more inner speech and less inner speech responded to our questionnaire. For example, participants with more inner speech were more likely to say that they get songs stuck in their head often, that they simulate both past and future conversations, that they rehearse the exact phrasing of a question they want to ask in a lecture before they ask it, and that “singing along” mentally feels like “regular thinking”. For the full set of questionnaire results, the reader is referred to Article IV and its accompanying appendices.

2.4.4. Conclusions & implications

In this article, we found that there are indeed behavioural differences between people who claim that their inner experience largely takes place in a verbal format and people who do not. These differences mainly manifested in verbal working memory performance and rhyme judgments, two tasks straightforwardly relying on verbal rehearsal and covert sound comparison. However, we did not find either the expected differences related to category processing or expected differences between the two groups in the task switching experiment. The latter is surprising given the robust evidence from dual-task interference studies that inner speech is usually involved in task switching (Baddeley et al., 2001; Emerson & Miyake, 2003; Miyake et al., 2004; Nedergaard et al., 2022). Overall, our experiments indicate that inner speech and verbal working memory are intimately linked but that the experience of phonologically specified, expanded inner speech is not necessary to either facilitate visual judgments or the self-regulatory behavioural control exercised in the task switching paradigm. Importantly, our two groups did not differ on any of the measured demographic characteristics (age, gender, education level, dyslexia, or first language).

It may be the case that the performance differences we found were driven by some factor other than how verbally represented participants' thoughts were. For example, participants reporting more inner speech could be generally more motivated or conscientious than participants with less inner speech. We believe that this explanation is unlikely, however, since participants reporting more inner speech were not across-the-board faster and more accurate which we would otherwise expect if a more general factor was driving the differences. For example, there were no differences between the two groups in responses to the no-rhyme pairs in the rhyme judgment experiment, responses to the orthographically similar set in the verbal working memory experiment, or any of the conditions in the task switching experiment. It is worth addressing that we found a relationship between reported inner speech and behaviour here which was absent in Article II where neither reported self-talk frequency nor self-talk efficacy were related to the magnitude of the verbal interference effect. Potential explanations for this discrepancy include: 1) that we did not have enough participants reporting extreme values on the self-talk efficacy or self-talk frequency scales in Article II to detect a relationship (e.g., only three and four people – from Experiment 1 and 2 respectively – reporting that they "never" talk to themselves while exercising), and 2) that the self-talk efficacy question ('what effect does self-talk usually have on your performance?') demanded both retrospection *and* inferring causation, factors which notoriously make self-report less reliable (Berger et al., 2016; Ericsson & Simon, 1980; Johansson et al., 2005; Nisbett & Wilson, 1977).

It is also important to note that the experience of inner speech could be a question of inclination rather than ability – participants reporting less inner speech may not habitually use inner speech to solve problems or engage in conversations with themselves, but that does not mean that they have no verbal working memory or that they are unable to repeat or generate words internally if asked to do so. The interesting question of what real-world consequences this inclination might have in terms of for example communicative abilities, mental health, and alternative strategies for problem-solving is for future studies to explore.

3. GENERAL DISCUSSION

Inner speech is a prominent part of most people's conscious experience, and so it is important to understand whether and how inner speech influences cognition and behaviour. The work presented in the present thesis explores such questions using several different approaches: dual-task interference, experience sampling, and individual differences. On the one hand, inner speech appears to be recruited for the control of both the body and the mind (Articles I, II, and III). On the other hand, people who experience little to no inner speech seem to be able to achieve a similar level of behavioural self-regulation through other means as measured in a task switching paradigm (Article IV) and as indicated in several studies included in our systematic review of verbal interference (Article I). In the following, I will discuss what we have learned about the nature and functions of inner speech from the studies included in the present thesis.

3.1. What have we learned about the nature of inner speech?

As discussed in the Introduction to the present thesis, inner speech can be characterised as internal experience of language with strong ties to verbal working memory. It can be recruited for a range of cognitive tasks, such as information storage, self-regulation, planning, motivation, and communication practice. In the Introduction, I adopted Alderson-Day and Fernyhough's intentionally broad formulation ('Inner speech can be defined as the subjective experience of language in the absence of overt and audible articulation') as my working definition. In the following, I will expand this tentative definition and discuss what I believe the four articles of this thesis contribute to our understanding of the nature of inner speech. In particular, I will address the previously discussed aspects of inner speech related to speaking and hearing, and how embodied or abstracted inner speech is.

Although the most immediately useful findings of our systematic review of verbal interference (Article I) concern the role of inner speech in various cognitive functions, there are other results that are highly relevant in the present context. The systematic review sheds light on a distinction between inner speaking and inner hearing through the effects of various interference task types. One of the interference task types we described, verbal shadowing, involves both listening to speech and producing speech while another, articulatory suppression, only involves producing speech and not listening to speech (or at least only involves listening to your own speech). Furthermore, it appears that both irrelevant speech presented auditorily without articulation (Gilhooly et al., 1999; Henson et al., 2003; Logie et al., 1994) and silent articulation (soundless mouthing) (Macken & Jones, 1995; Murray, 1965; Saito, 1993) disrupt normal functioning of verbal working memory. Therefore, it seems that an account of inner speech purely focused on auditory imagery is unlikely to be adequate, as is a purely motor imagery account. On the one hand, this fits well with real-life experience where we never produce speech without also hearing it. If inner speech is internalised outer speech, as many theories hold, then it should rely on *both* auditory imagery and articulatory-motor imagery (see especially the motor simulation view). On the other hand, Descriptive Experience Sampling research suggests that we should distinguish between inner speaking and inner hearing, as do prominent theories of overt speech production and comprehension (Galantucci et al., 2006; Lotto et al., 2009). This may be a subtle – but relevant – phenomenological difference to discuss when people report hearing other people's voices and simulating conversations (Feigenbaum, 2009; Gregory, 2016; Honeycutt, 2020). None of the dual-task experiments reviewed in Article I focused on simulated conversations or hypothesised that the mental imagery of other people's speech would be important. If future dual-task studies were to focus on these, we might see interesting differences between the effects of the different interference types (targeting articulation or audition).

The studies in this thesis also contribute to our understanding of the relationship between inner speech and verbal working memory. One of the more striking differences in our anendophasia study (Article IV) was in the verbal working memory experiment, where participants who reported more inner speech were better at immediate serial recall of words. Because propensity to engage in inner speech was predictive of verbal working memory capacity, our study corroborates the idea that inner speech and verbal working memory are indeed connected – as do the verbal interference results from Articles I and II where occupying the phonological loop in many cases disrupted hypothesised inner speech functions. The correlation between inner speech and verbal working memory in our anendophasia study is also especially interesting because the inner speech scores were entirely constructed from participants' self-report about their inner speech experience. We could otherwise imagine that self-reported inner speech use and verbal working memory capacity were independent of each other, but this does not seem to be the case. Our results provide some nuance to findings that spontaneous inner speech and lab-elicited inner speech (often essentially verbal working memory tasks) have different neural substrates (Hurlburt et al., 2016). Spontaneous inner speech may be generally more condensed and abstracted than the kind of explicit inner speech needed in verbal working memory tasks, but our results suggest that these two measures are positively correlated within individuals. The correlation between reported inner speech and verbal working memory is also interesting given Descriptive Experience Sampling findings that people are often mistaken in their reports about inner speech (Hurlburt et al., 2013; Hurlburt & Heavey, 2015). Our anendophasia results suggest that they might not be *that* mistaken (although see section 3.4. 'General limitations of the present studies' below for further discussion).

Both the findings from Article I and Article IV are also relevant for the debate on how abstracted or embodied inner speech is. As previously discussed, inner speech can be thought of as condensed/abstracted or expanded/embodied in terms of its phonology, syntax, and semantics. Some theories claim that the extent to which inner speech is experienced as expanded depends on situational factors (such as cognitive demand, stress, or social isolation; Brinthaupt, 2019; Fernyhough, 2004). Article I indicated that interfering with the phonological loop can disrupt a number of inner speech functions, providing support for a relatively embodied conception of inner speech. If inner speech were entirely abstracted away from articulatory and auditory processes, repeating "the" out loud should not interfere with inner speech. The results from Article IV indicate that the condensed-expanded dimension may also vary across individuals in addition to across situations. Participants who reported less inner speech performed less well on verbal working memory and rhyme judgment tasks which require a high level of phonological specification, but they did not show different performance from participants reporting more inner speech on tasks that may be completed using less phonologically specified inner speech. For example, it could be possible to keep track of plus or minus operations without needing to say or hear /plʌs/ and /maɪnəs/ through the "soundless" words/concepts PLUS and MINUS. Two interpretations are possible: 1) individuals with anendophasia do not experience inner speech at all at any level of abstraction, or 2) individuals with anendophasia experience inner speech at a level of abstraction where phonological information is not specified. If the former is true, then inner speech is presumably not crucially important for any aspect of cognition given that these individuals do not generally "stand out". If the latter is true, then we have learned something important about expanded nature of inner speech and its relation to verbal working memory. It should be possible to test whether individuals with anendophasia are still influenced by the words of natural language without phonological specification, for example by examining categorical perception of colour in such individuals. Some theorists have suggested that the "unsymbolised thinking" that DES participants sometimes report is the most condensed form of inner speech (Vicente & Martínez-Manrique, 2016), but I do not agree with this characterisation for the following reasons: first, unsymbolised thinking as described in DES studies does not have the temporal properties that even condensed inner speech has (rhythm, pace, etc.) (Hurlburt & Akhter, 2006, 2008), and second, it is not clear what properties of unsymbolised thought that are sufficiently language-like to justify subsuming under "inner speech" rather than under "thought".

In summary, the four studies included in the present thesis contribute to our understanding of the nature of inner speech in that they 1) provide evidence that inner speech involves *both* motor imagery and auditory imagery, 2) indicate support for the flexible abstraction of inner speech, and 3) underline the strong connections between inner speech and verbal working memory.

3.2. What have we learned about the functions of inner speech?

It appears that inner speech is not just a by-product of speech and communication planning – people also use it to regulate their own behaviour. This self-regulatory function plays a central role in Vygotskian accounts of inner speech. Two of the studies included in the present thesis are especially relevant for this perspective: the cycling experiments (Article II) and the online sustained attention experiment (Article III). The first study concerns the role of inner speech in the control of the body (physical endurance) while the second concerns the role of inner speech in the control of the mind (sustained attention).

Research in cognitive science has not focused very much on the effects of inner speech on controlling the body, but, fortunately, sport psychology suffers from no such oversight. There is a rich literature concerning the antecedents and consequences of both spontaneous and goal-directed self-talk (Dickens et al., 2018; Latinjak et al., 2019; Tod et al., 2011; Van Raalte et al., 2016). This literature places emphasis on intervention studies where athletes are trained to talk to themselves in specific ways, and it appears that self-talk trained in this manner does have beneficial effects on performance, especially if tailored to the specific demands of a given sport (Hardy et al., 2015; Theodorakis et al., 2000; Zourbanos et al., 2013). However, the intervention studies in many cases suffer from low sample sizes and lack of active control conditions (see Article II for further discussion). A few studies have used interference tasks (Biese et al., 2019; Blakely et al., 2016; Darling & Helton, 2014; Green & Helton, 2011; Stets et al., 2020; Talarico et al., 2017), but not controlled in a way that would allow conclusions about inner speech involvement, e.g., the interference tasks were not matched in terms of difficulty or presence or absence of verbality. Our study presented in Article II does exactly this and provides direct evidence that inner speech under normal circumstances helps control endurance performance. What is interesting in this context is *how* this happens. In sport psychology, there are theories that self-talk helps endurance indirectly by influencing *perceived exertion* (Aitchison et al., 2013). I believe an addition to this could be that inner speech helps with impulse control (Mischel et al.,

1996) – in the case of physical endurance, the body's prepotent response would be to quit, and people are less able to inhibit this response when distracted from talking to themselves (see also Tullett & Inzlicht, 2010).

In a similar vein, inner speech could also be related to response inhibition in the online sustained attention experiment presented in Article III. In this experiment, the task was so boring that the mind's prepotent response would be to wander. Similar to the inhibition of a quitting response in the physical domain, we hypothesise that participants use inner speech to prevent themselves from giving in to this temptation in the cognitive domain. Like the cycling experiments in Article II, this is a straightforward case of where inner speech functions the same way that someone else's speech would. There is a natural line from a caregiver telling a child to pay attention or not to give up to an adult telling themselves the same things (Vygotsky, 1962). Some of the studies reviewed in Article I also showed that verbal interference was associated with a lack of inhibitory control (Dunbar & Sussman, 1995; Tullett & Inzlicht, 2010; Wallace et al., 2017). Both Article II and Article III, of course, have their limitations. In the cycling experiments, participants' self-reported inner speech frequency and efficacy were unrelated to how much they were affected by verbal interference. This at least suggests that we should not put too much trust in self-reports about inner speech effects and consequences. In the online sustained attention study, the differences in reaction times were quite small, and we had few trials per participant to draw conclusions from. Nevertheless, these two studies do, in my view, contribute novel insights into the mechanisms of how inner speech can influence behaviour and cognition.

3.3. What are the limits of inner speech functions?

Judging by the articles included in the present thesis, it appears that inner speech can function as a tool for attentional and behavioural control. However, Articles I and IV also highlight some limits to inner speech functions.

Article I found that verbal interference did not have specific disruptive effects on some cognitive tasks with hypothesised verbal involvement. In particular, we did not find specific disruptive effects of verbal interference for primary tasks relying on visual processing (the categories we labelled visual change, visuospatial integration and wayfinding, and reasoning using non-verbal materials) or theory of mind. For visual processing, the hypothesis put forward in the literature was that inner speech plays a role either by providing a common format for integrating information across different modalities (Hermer-Vazquez et al., 1999) or as a part of dual (both visual and verbal) encoding (Paivio, 1991). For theory of mind, inner speech was similarly hypothesised to provide a format for representing mental states (de Villiers, 2007; de Villiers & de Villiers, 2000). In both these cases, inner speech was thus hypothesised to play a role as a structuring tool rather than a rehearsal medium. The fact that verbal interference did not appear to affect these processes indicates either that inner speech is simply not involved or that dual-task interference is too superficial a tool to target such structuring inner speech functions. Given that dual-task interference sometimes did appear to affect other structuring inner speech functions like categorization (Lupyan, 2009; Maddox et al., 2004; Minda et al., 2008; Souza & Skóra, 2017; Zeithamova & Maddox, 2007), the former explanation is perhaps more likely.

While there is ample evidence for a role of inner speech in behavioural control – both in the articles included in the present thesis and elsewhere – the task switching experiment from Article IV provides important nuance. In that experiment, there was no difference between the group with more inner speech and the group with less inner speech in terms of task switching speed and accuracy. This could suggest that the self-regulatory functions of inner speech can also be achieved through other means, which in turn prompts new questions: What is it about inner speech that makes it such a common tool for self-regulation? Can we gain insight into how this self-regulatory function is fulfilled by looking at alternative strategies? We can get some clues from the anendophasia study by looking at what participants claimed to be doing to remember the appropriate action (adding or subtracting). All their non-language-based strategies were about creating either some striking mnemonic (visualising a cartoon character giving thumbs-up or thumbs-down) or an external prompt (tapping one finger to mean subtracting and another to mean adding). If the task rules had been more complicated than just two alternating options, we might have seen inner speech play a more important role as language can represent any number of task rules and is not limited by e.g., the number of fingers to tap. These questions will be for future research to explore.

As previously discussed, it seems likely that people with anendophasia do in fact experience inner speech, but with a lower amount of articulatory/auditory imagery. From a motor simulation point of view, we can think of their inner speech as abandoned at an earlier planning/simulation stage, and thus experienced as less fully specified in terms of phonology and articulation. Individual differences in imagery vividness are well-described and appear to correlate across sensory modalities (Dawes et al., 2020; Roebuck & Lupyan, 2020). Indeed, the participants with anendophasia in Article IV also reported less vivid auditory imagery in general than the control group (e.g., songs stuck in their heads less often). It seems likely that the lack of the experience of phonologically specified inner speech has downstream effects on what people with anendophasia use inner speech for. For example, they should be less likely to use it for specifically language-based situations where both "speaking" and "hearing" simulated utterances are essential. Indeed, one of the clearest differences between the two groups in our inner speech questionnaire in Article IV was that participants with less inner speech were much less likely to report that they habitually simulate past and future conversations. The fact that an inner monologue or dialogue may not come naturally to people with anendophasia should be taken into account in future research. Further explorations of how such individuals do think and solve problems will be important for our understanding of the variety of different cognitive task solving strategies available to humans. This will be relevant for interventions in sport (Hatzigeorgiadis et al., 2011; Tod et al., 2011), therapy (Hollon & Beck, 2013), and education (Deniz, 2009).

3.4. General limitations of the present studies

The theoretical and methodological limitations of each of the four studies are considered in detail in sections covering the relevant articles as well as in the articles themselves, but it is worth discussing some more overarching issues.

First, large parts of the present thesis rely on the dual-task interference method. This method is a blunt instrument and can only indirectly be used to claim anything specific about *how* inner speech benefits a certain task even if there is a specific disruptive effect of verbal interference. To make such claims, it will for example be necessary to combine interventions that up-regulate the use of specific inner speech statements with verbal interference. If saying 'you can do it!' to yourself is more effective than saying 'you suck!' to yourself, then verbal interference should have a more disruptive effect in the former case. Such interventions are well known in sport psychology, but they have not been combined with verbal interference and an appropriate control. Relatedly, it could also be a problem that our studies assume a modular construal of the mind where all thought in a verbal format (condensed or expanded inner speech) can be interrupted by verbal interference. There are reasons to believe that the way inner speech works might not be so straightforward. For example, DES experiments have shown that people experience simultaneously saying one thing with their "outer" speech and another thing with their "inner" speech (Hurlburt & Heavey, 2018) – not just different versions of the same sentiment but also radically different meanings. This suggests that inner speech and outer speech may in some cases function independently. This is of course a problem when outer speech interference is used to target inner speech. However, this does not appear to be common, and the dual-task method has been shown to be reliable in so many cases (see Baddeley & Hitch, 2019, for an overview) that the current results plausibly capture relevant aspects of the phenomenon of interest nonetheless.

Second, there is the problem of relying on self-report to study inner speech. I discussed this in section 1.2. ('How can we know that people use inner speech?'), but it is worth re-examining in light of the articles constituting the present thesis. In particular, the contrast between the self-report/behaviour relationship in Article II and the self-report/behaviour relationship in Article IV deserves attention. In Article II, participants were less able to increase their endurance performance when under verbal interference than when under visuospatial interference. However, this interference effect did not vary with either how much participants reported that they usually talk to themselves while exercising or with how much they think it helps them to talk to themselves while exercising. In Article IV, participants' self-reported inner speech use was related to their verbal working memory and picture-naming performance. It is a well-known finding from metacognition research that self-report might not be so reliable when people are asked to make inferences about their experience rather than simply describe it ('how did Y affect you?', 'why did you choose X?') (Berger et al., 2016; Johansson et al., 2005; Nisbett & Wilson, 1977). This finding corresponds well to our results in Article II, especially the absence of an effect of how much participants reported that talking to themselves usually helps them. The inner speech scores in Article IV were based on participants describing their experience, which is generally more reliable than asking people to infer causation (Ericsson & Simon, 1980; Petitmengin et al., 2013).

3.5. Future studies

In the studies included in the present thesis, we focused on dual-task interference, experience sampling, and individual differences approaches. Further investigations growing directly out of those studies are detailed in the respective sections, but there are potential fruitful intersections as well. For example, it would be interesting to conduct the kind of sustained attention study detailed in Article III with participants who report no habitual use of inner speech to see if their experience samples are consistent with such general reports, and if they are, what they do instead to maintain attention. In addition, the different effects of two types of interference tasks in Article II could be used to inform or challenge the findings from Article I, especially the studies that used memory-based interference which our cycling experiments indicated has a relatively weak interference effect. The question of different inner speech strategies is a common thread in all four studies – in Article I, some of the studies specifically investigated training in different strategies and found different effects of verbal interference, while in Articles III and IV we specifically asked them what kind of strategy they had used to solve the tasks. The combination of training different strategies (up-regulating specific inner speech use) and interfering with inner speech (down-regulating) is also a promising avenue for future explorations.

There are of course other methodological avenues to explore such as developmental, neurological, and physiological correlates of inner speech. These three correlates are particularly interesting because they help us triangulate the problem of having to rely on self-report. Developmental studies provide a good window into inner speech because children engage in private, overt speech which can be independently observed (caveats about equating inner speech and private speech aside). While private speech in children has been the focus of much research already, individual differences in the use of private speech in adults have not received much attention. It would be interesting to study different inner speech-based problem-solving strategies – if the Vygotskian conception of the developmental trajectory is correct, we would expect the strategies used in private speech to continue into inner speech in adulthood. Taking such an individual differences approach would also illuminate the possible developmental origins of inner speech variability – for individuals experiencing less inner speech, it might be the case that they have received less verbally shaped external regulation or that the internalisation process for some reason resulted in a more condensed and abstracted version of inner speech.

Neuroimaging may be used to further elaborate how overt and covert speech are related as well as the relative contributions of inner speaking/articulatory-motor imagery and inner hearing/auditory imagery. Similarly, other physiological approaches let us both study the potential consequences of inner speech – altered heart rate, galvanic skin conductance, cortisol, etc. – and further explore motor simulation views on inner speech for example with electromyography. As these methods are developed further, we could get closer to objective measures of inner speech in adults. Electromyography measures of inner speech assume that inner speech is relatively embodied because it measures minuscule movements of the articulatory muscles (Garrity, 1977; Moffatt et al., 2020; Nalborczyk et al., 2020), but this method could also be used to further explore the flexibly abstracted nature of inner speech, e.g., whether it is more like overt speech in cognitively demanding situations like preparing for public speaking or solving puzzles under time pressure.

4. CONCLUSION

In this thesis, I have described my research into the phenomenon of inner speech, what it could be, how it might be measured, and its potential functions. I have presented four articles – a systematic review of verbal interference as a method to investigate the role of language in cognition (I), a dual-task interference study of physical endurance (II), an experience sampling study of the role of inner speech in sustained attention (III), and a behavioural individual differences study of people with little to no inner speech (anendophasia) (IV). These studies underline the intimate connection between inner speech and verbal working memory, provide much-needed nuance to the verbal interference method and what can be concluded from it, explore the use of inner speech in physical and mental endurance, and present intriguing evidence that differences in inner speech experience are both largely masked and have measurable behavioural consequences. Research into both the nature and functions of inner speech continues to hold much potential for future methodological and theoretical advances. The work detailed above provides a fuller picture of the role of inner speech in cognition and contributes to stronger, more nuanced theories of how inner speech can function as an important tool for self-regulation and behavioural control.

5. REFERENCES

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*(7), 1059–1068. https://doi.org/10.3758/BF03205520

Agnati, L. F., Barlow, P., Ghidoni, R., Borroto-Escuela, D. O., Guidolin, D., & Fuxe, K. (2012). Possible genetic and epigenetic links between human inner speech, schizophrenia and altruism. *Brain Research*, *1476*, 38–57. https://doi.org/10.1016/j.brainres.2012.02.074

Aitchison, C., Turner, L. A., Ansley, L., Thompson, K. G., Micklewright, D., & Gibson, A. S. C. (2013). Inner dialogue and its relationship to perceived exertion during different running intensities. *Perceptual and Motor Skills*, *117*(1), 11–30.

Alderson-Day, B., & Fernyhough, C. (2014). More than one voice: Investigating the phenomenological properties of inner speech requires a variety of methods. *Consciousness and Cognition*, *24*, 113–114. https://doi.org/10.1016/j.concog.2013.12.012

Alderson-Day, B., & Fernyhough, C. (2015a). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00517

Alderson-Day, B., & Fernyhough, C. (2015b). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

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. https://doi.org/10.1016/j.concog.2018.07.001

Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi

Arabian and British samples. *Developmental Psychology*, *42*(1), 117–131. https://doi.org/10.1037/0012-1649.42.1.117

Anderson, R. E. (1982). Speech imagery is not always faster than visual imagery. *Memory & Cognition*, *10*(4), 371–380. https://doi.org/10.3758/BF03202429

Astington, J. W., & Baird, J. A. (2005). *Why language matters for theory of mind*. Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641–657. https://doi.org/10.1037/0096-3445.130.4.641

Baddeley, A. D. (1966a). Short-term Memory for Word Sequences as a Function of Acoustic, Semantic and Formal Similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362–365. https://doi.org/10.1080/14640746608400055

Baddeley, A. D. (1966b). The Influence of Acoustic and Semantic Similarity on Long-term Memory for Word Sequences. *Quarterly Journal of Experimental Psychology*, *18*(4), 302–309. https://doi.org/10.1080/14640746608400047

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2007). *Working Memory, Thought, and Action*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. D. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641–657. https://doi.org/10.1037/0096-3445.130.4.641

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex*, *112*, 91–106. https://doi.org/10.1016/j.cortex.2018.05.015

Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology*, *60*(4), 497–504. https://doi.org/10.1080/17470210601147572

Baddeley, A. D., & Lewis, V. (1981). Inner active processes in reading: The inner voice, the inner ear, and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 107–129). Lawrence Erlbaum.

Baddeley, A. D., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, *24*(4), 490–502. https://doi.org/10.1016/0749-596X(85)90041-5

Barwood, M. J., Corbett, J., Wagstaff, C. R., McVeigh, D., & Thelwell, R. C. (2015). Improvement of 10-km time-trial cycling with motivational self-talk compared with neutral self-talk. *International Journal of Sports Physiology and Performance*, *10*(2), 166–171.

Basho, S., Palmer, E. D., Rubio, M. A., Wulfeck, B., & Müller, R.-A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: Paced production and overt speech. *Neuropsychologia*, *45*(8), 1697–1706. https://doi.org/10.1016/j.neuropsychologia.2007.01.007

Bek, J., Blades, M., Siegal, M., & Varley, R. (2013). Dual-Task Interference in Spatial Reorientation: Linguistic and Nonlinguistic Factors. *Spatial Cognition & Computation*, *13*(1), 26–49. https://doi.org/10.1080/13875868.2011.590622

Berger, C. C., Dennehy, T. C., Bargh, J. A., & Morsella, E. (2016). Nisbett and Wilson (1977) revisited: The little that we can know and can tell. *Social Cognition*, *34*(3), 167.

Berk, L. E. (1992). Children's private speech: An overview of theory and the status of research. In R. M. Diaz & L. E. Berk (Eds.), *Private Speech: From Social Interaction To Self-regulation* (pp. 17–54). Psychology Press.

Berk, L. E. (1994). Why Children Talk to Themselves. *Scientific American*, *271*(5), 78–83. https://doi.org/10.1038/scientificamerican1194-78

Berk, L. E., & Garvin, R. A. (1984). Development of private speech among low-income Appalachian children. *Developmental Psychology*, *20*(2), 271–286. https://doi.org/10.1037/0012-1649.20.2.271

Bermúdez, J. L. (2018). Inner Speech, Determinacy, and Thinking Consciously about Thoughts. In *Inner Speech: New Voices* (Vol. 1, pp. 199–220). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0008

Besner, D., Davies, J., & Daniels, S. (1981). Reading for Meaning: The Effects of Concurrent Articulation. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 415–437. https://doi.org/10.1080/14640748108400801

Biese, K. M., Pietrosimone, L. E., Andrejchak, M., Lynall, R. C., Wikstrom, E. A., & Padua, D. A. (2019). Preliminary Investigation on the Effect of Cognition on Jump-Landing Performance Using a Clinically Relevant Setup. *Measurement in Physical Education and Exercise Science*, *23*(1), 78–88. https://doi.org/10.1080/1091367X.2018.1518875

Bivens, J. A., & Berk, L. E. (1990). A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly (1982-)*, 443–463.

Blakely, M. J., Kemp, S., & Helton, W. S. (2016). Volitional Running and Tone Counting: The Impact of Cognitive Load on Running Over Natural Terrain. *IIE Transactions on Occupational Ergonomics and Human Factors*, *4*(2–3), 104–114. https://doi.org/10.1080/21577323.2015.1055864

Blonskii, P. P. (1964). Memory and Thought. In *Selected Works in Psychology*. Prosveshchenie Press.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 157–165. https://doi.org/10.1037/0278-7393.18.1.157

Brick, N. E., MacIntyre, T. E., & Campbell, M. J. (2016). Thinking and Action: A Cognitive Perspective on Self-Regulation during Endurance Performance. *Frontiers in Physiology*, *7*. https://doi.org/10.3389/fphys.2016.00159

Brinthaupt, T. M. (2019). Individual Differences in Self-Talk Frequency: Social Isolation and Cognitive Disruption. *Frontiers in Psychology*, *10*, 1088. https://doi.org/10.3389/fpsyg.2019.01088

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. https://doi.org/10.1080/00223890802484498

Caffò, A. O., Picucci, L., Di Masi, M. N., & Bosco, A. (2011). Working memory components and virtual reorientation: A dual-task study. In E. S. Levin (Ed.), *Working memory: Capacity, developments and improvement techniques* (pp. 249–266). Nova Science Publisher.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674. https://doi.org/10.1017/S0140525X02000122

Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: Effect of cue transparency on children's cognitive flexibility. *Developmental Psychology*, *45*(3), 782–797. https://doi.org/10.1037/a0015409

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, *28*(9), 868–879. https://doi.org/10.1002/hbm.20315

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought* (1st ed., pp. 162–183). Cambridge University Press. https://doi.org/10.1017/CBO9780511597909.011

Clark, A. (2008). 3 Material Symbols. In A. Clark, *Supersizing the Mind* (1st ed., pp. 44–60). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195333213.003.0003

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, *55*, 75–84.

Conrad, R., & Hull, A. J. (1964). INFORMATION, ACOUSTIC CONFUSION AND MEMORY SPAN. *British Journal of Psychology*, *55*(4), 429–432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 162–175. https://doi.org/10.1037/a0021321

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

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sample method. *Journal of Nervous and Mental Disease*, *175*(9), 526–536.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*(3), 281–294. https://doi.org/10.1007/BF02138940

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

Darling, K. A., & Helton, W. S. (2014). Dual-task interference between climbing and a simulated communication task. *Experimental Brain Research*, *232*(4), 1367–1377. https://doi.org/10.1007/s00221-014-3855-7

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.

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua*, *117*(11), 1858–1878.

de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Psychology Press.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, *103*(38), 14240–14245. https://doi.org/10.1073/pnas.0606302103

Dell, G. S., & Repka, R. J. (1992). Errors in Inner Speech. In B. J. Baars (Ed.), *Experimental Slips and Human Error* (pp. 237–262). Springer US. https://doi.org/10.1007/978-1-4899-1164-3_10

Deniz, C. (2009). Early childhood teachers' awareness, beliefs, and practices toward children's private speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 236–246). Cambridge University Press.

Diaz, R. M., & Berk, L. E. (Eds.). (1992). *Private Speech* (0 ed.). Psychology Press. https://doi.org/10.4324/9781315807270

Dickens, Y. L., Van Raalte, J., & Hurlburt, R. T. (2018). On Investigating Self-Talk: A Descriptive Experience Sampling Study of Inner Experience During Golf Performance. *The Sport Psychologist*, *32*(1), 66–73. https://doi.org/10.1123/tsp.2016-0073

Dunbar, K., & Sussman, D. (1995). Toward a Cognitive Account of Frontal Lobe Function: Simulating Frontal Lobe Deficits in Normal Subjects. *Annals of the New York Academy of*

Sciences, *769*(1 Structure and), 289–304. https://doi.org/10.1111/j.1749-6632.1995.tb38146.x

Dungan, J., & Saxe, R. (2012). Matched False-Belief Performance During Verbal and Nonverbal Interference. *Cognitive Science*, *36*(6), 1148–1156. https://doi.org/10.1111/j.1551-6709.2012.01248.x

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. https://doi.org/10.1016/S0749-596X(02)00511-9

Erdmann, J. E. (1851). *Psychologische Briefe*. Reichardt.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Feigenbaum, P. (2009). Development of Communicative Competence through Private and Inner Speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation* (1st ed., pp. 105–120). Cambridge University Press. https://doi.org/10.1017/CBO9780511581533.009

Fernyhough, C. (1996). The dialogic mind: A dialogic approach to the higher mental functions. *New Ideas in Psychology*, *14*(1), 47–62. https://doi.org/10.1016/0732-118X(95)00024-B

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

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Forgeot d'Arc, B., & Ramus, F. (2011). Belief attribution despite verbal interference. *Quarterly Journal of Experimental Psychology*, *64*(5), 975–990. https://doi.org/10.1080/17470218.2010.524413

Frankish, K. (2018). Inner speech and outer thought. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: New Voices*. Oxford University Press.

Frings, M., Dimitrova, A., Schorn, C. F., Elles, H.-G., Hein-Kropp, C., Gizewski, E. R., Diener, H. C., & Timmann, D. (2006). Cerebellar involvement in verb generation: An fMRI study. *Neuroscience Letters*, *409*(1), 19–23. https://doi.org/10.1016/j.neulet.2006.08.058

Gagne, D. L., & Coppola, M. (2017). Visible Social Interactions Do Not Support the Development of False Belief Understanding in the Absence of Linguistic Input: Evidence from Deaf Adult Homesigners. *Frontiers in Psychology*, *8*, 837. https://doi.org/10.3389/fpsyg.2017.00837

Galanis, E., Hatzigeorgiadis, A., Comoutos, N., Papaioannou, A., Morres, I. D., & Theodorakis, Y. (2021). Effects of a strategic self-talk intervention on attention functions. *International Journal of Sport and Exercise Psychology*, 1–15. https://doi.org/10.1080/1612197X.2021.1963304

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. https://doi.org/10.3758/BF03193857

Garrity, L. I. (1977). Electromyography: A review of the current status of subvocal speech research. *Memory & Cognition*, *5*(6), 615–622. https://doi.org/10.3758/BF03197407

Gauker, C. (2018). Inner Speech as the Internalization of Outer Speech. In *Inner Speech: New Voices* (Vol. 1, pp. 53–77). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0003

Geiger, L. (1868). *Ursprung und Entwickelung der menschlichen Sprache und Vernunft.* Verlag der J. G. Cotta'schen Buchhandlung.

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, *134*(10), 3071–3082. https://doi.org/10.1093/brain/awr232

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic Reasoning Tasks, Working Memory, and Skill. *European Journal of Cognitive Psychology*, *11*(4), 473–498. https://doi.org/10.1080/095414499382264

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Attention & Performance XVIII: Control of Cognitive Processes*. MIT Press.

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.-P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & 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

Green, A. L., & Helton, W. S. (2011). Dual-task performance during a climbing traverse. *Experimental Brain Research*, *215*(3–4), 307–313. https://doi.org/10.1007/s00221-011-2898-2

Gregory, D. (2016). Inner Speech, Imagined Speech, and Auditory Verbal Hallucinations. *Review of Philosophy and Psychology*, *7*(3), 653–673. https://doi.org/10.1007/s13164-015-0274-z

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*(1), 1–19. https://doi.org/10.1002/1097-0193(200101)12:1<1::AID-HBM10>3.0.CO;2-V

Hammer, K. (2019). Bilingual cogito: Inner speech in acculturated bilinguals. *International Journal of Bilingual Education and Bilingualism*, *22*(5), 576–592. https://doi.org/10.1080/13670050.2017.1285862

Hardy, J., Begley, K., & Blanchfield, A. W. (2015). It's Good But it's Not Right: Instructional Self-Talk and Skilled Performance. *Journal of Applied Sport Psychology*, *27*(2), 132–139. https://doi.org/10.1080/10413200.2014.959624

Hatzigeorgiadis, A., & Galanis, E. (2017). Self-talk effectiveness and attention. *Current Opinion in Psychology*, *16*, 138–142. https://doi.org/10.1016/j.copsyc.2017.05.014

Hatzigeorgiadis, A., Theodorakis, Y., & Zourbanos, N. (2004). Self-Talk in the Swimming Pool: The Effects of Self-Talk on Thought Content and Performance on Water-Polo Tasks. *Journal of Applied Sport Psychology*, *16*(2), 138–150. https://doi.org/10.1080/10413200490437886

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., & Theodorakis, Y. (2011). Self-Talk and Sports Performance: A Meta-Analysis. *Perspectives on Psychological Science*, *6*(4), 348–356. https://doi.org/10.1177/1745691611413136

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, *17*(3), 798–810. https://doi.org/10.1016/j.concog.2007.12.006

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective Interference with Verbal Short-Term Memory for Serial Order Information: A New Paradigm and Tests of a Timing-Signal Hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1307–1334. https://doi.org/10.1080/02724980244000747

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of Flexibility in Human Cognition: Dual-Task Studies of Space and Language. *Cognitive Psychology*, *39*(1), 3–36. https://doi.org/10.1006/cogp.1998.0713

Hinwar, R. P., & Lambert, A. J. (2021). Anauralia: The Silent Mind and Its Association With Aphantasia. *Frontiers in Psychology*, *12*, 744213. https://doi.org/10.3389/fpsyg.2021.744213

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination. *Memory & Cognition*, *23*(2), 147–154. https://doi.org/10.3758/BF03197217

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Heffernan, T. M. (1991). Speech, "inner speech," and the development of short-term memory: Effects of picture-labeling on recall. *Journal of Experimental Child Psychology*, *51*(2), 220–234. https://doi.org/10.1016/0022-0965(91)90033-O

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., McMillon, G., Kolchugina, G., Black, J. M., Faizi, A., & others. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*, *104*(10), 4234–4239.

Hollon, S. D., & Beck, A. T. (2013). Cognitive and cognitive behavioral therapies. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavioral change* (6th ed., pp. 447–492). Wiley.

Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research*, *4*(4), 383–395. https://doi.org/10.1007/BF01178214

Honeycutt, J. M. (2020). Imagined Interactions and Inner Speech. *Imagination, Cognition and Personality*, *39*(4), 386–396. https://doi.org/10.1177/0276236619869163

Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality*, *54*(1), 262–293. https://doi.org/10.1111/j.1467-6494.1986.tb00395.x

Huang, J., Carr, T. H., & Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping*, *15*(1), 39–53. https://doi.org/10.1002/hbm.1060

Hupbach, A., Hardt, O., Nadel, L., & Bohbot, V. D. (2007). Spatial Reorientation: Effects of Verbal and Spatial Shadowing. *Spatial Cognition & Computation*, *7*(2), 213–226. https://doi.org/10.1080/13875860701418206

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., 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. https://doi.org/10.1371/journal.pone.0147932

Hurlburt, R. T., & Heavey, C. L. (2006). *Exploring Inner Experience: The descriptive experience sampling method* (Vol. 64). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.64

Hurlburt, R. T., & Heavey, C. L. (2015). Investigating pristine inner experience: Implications for experience sampling and questionnaires. *Consciousness and Cognition*, *31*, 148–159. https://doi.org/10.1016/j.concog.2014.11.002

Hurlburt, R. T., & Heavey, C. L. (2018). Inner Speaking as Pristine Inner Experience. In *Inner Speech: New Voices* (Vol. 1, pp. 168–196). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0007

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., Heavey, C. L., Lapping-Carr, L., Krumm, A. E., Moynihan, S. A., Kaneshiro, C., Brouwers, V. P., Turner, D. K., & Kelsey, J. M. (2022). Measuring the Frequency of Inner-Experience Characteristics. *Perspectives on Psychological Science*, *17*(2), 559–571. https://doi.org/10.1177/1745691621990379

Hyland-Monks, R., Cronin, L., McNaughton, L., & Marchant, D. (2018). The role of executive function in the self-regulation of endurance performance: A critical review. In *Progress in Brain Research* (Vol. 240, pp. 353–370). Elsevier. https://doi.org/10.1016/bs.pbr.2018.09.011

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

Jacobson, E. (1930). Electrical measurements of neuromuscular states during mental activities VII. Imagination, recollection and abstract thinking involving the speech musculature. *American Journal of Physiology*, *97*, 200–209.

Jacobson, E. (1932). Electrophysiology of mental activities. *The American Journal of Psychology*, *44*, 677–694.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202. https://doi.org/10.1017/S0140525X00034026

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science*, *310*(5745), 116–119. https://doi.org/10.1126/science.1111709

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology*, *60*(4), 505–511. https://doi.org/10.1080/17470210601147598

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 656–674.

Keeney, T. J., Cannizzo, S. R., & Flavell, J. H. (1967). Spontaneous and Induced Verbal Rehearsal in a Recall Task. *Child Development*, *38*(4), 953. https://doi.org/10.2307/1127095

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.

Kirschenbaum, D. S. (1987). Self-regulation of sport performance: *Medicine & Science in Sports & Exercise*, *19*(Supplement), S114–S120. https://doi.org/10.1249/00005768-198710001-00003

Korba, R. J. (1990). The Rate of Inner Speech. *Perceptual and Motor Skills*, *71*(3), 1043–1052. https://doi.org/10.2466/pms.1990.71.3.1043

Kray, J., Eber, J., & Karbach, J. (2008). Verbal self-instructions in task switching: A compensatory tool for action-control deficits in childhood and old age? *Developmental Science*, *11*(2), 223–236. https://doi.org/10.1111/j.1467-7687.2008.00673.x

Langland-Hassan, P. (2018). From Introspection to Essence: The Auditory Nature of Inner Speech. In P. Langland-Hassan & A. Vicente (Eds.), *Inner Speech: New Voices* (pp. 78–104). Oxford University Press.

Larrain, A., & Haye, A. (2012). The discursive nature of inner speech. *Theory & Psychology*, *22*(1), 3–22. https://doi.org/10.1177/0959354311423864

Larsen, J. D., & Baddeley, A. D. (2003). Disruption of Verbal Stm by Irrelevant Speech, Articulatory Suppression, and Manual Tapping: Do they have a Common Source? *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1249–1268. https://doi.org/10.1080/02724980244000765

Larsen, S. F., Schrauf, R. W., Fromholt, P., & Rubin, D. C. (2002). Inner speech and bilingual autobiographical memory: A Polish-Danish cross-cultural study. *Memory*, *10*(1), 45–54. https://doi.org/10.1080/09658210143000218

Latinjak, A. T., Hatzigeorgiadis, A., Comoutos, N., & Hardy, J. (2019). Speaking clearly . . . 10 years on: The case for an integrative perspective of self-talk in sport. *Sport, Exercise, and Performance Psychology*, *8*(4), 353–367. https://doi.org/10.1037/spy0000160

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83–122). Oxford University Press.

Liefooghe, B., Vandierendonck, A., Muyllaert, I., Verbruggen, F., & Vanneste, W. (2005). The phonological loop in task alternation and task repetition. *Memory*, *13*(5), 550–560. https://doi.org/10.1080/09658210444000250

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert Speech Behavior during a Silent Language Recitation Task. *Perceptual and Motor Skills*, *83*(3_suppl), 1355–1362. https://doi.org/10.2466/pms.1996.83.3f.1355

Loevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciu, M., & Perrone-Bertolotti, M. (2018). A Cognitive Neuroscience View of Inner Language: To Predict and to Hear, See, Feel. In P. Langland-Hassan & A. Vicente (Eds.), *Inner Speech: New Voices*. Oxford University Press.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, *22*(4), 395–410. https://doi.org/10.3758/BF03200866

Logie, R. H., Sala, S. D., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305–321. https://doi.org/10.3758/BF03213295

Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development*, *74*(4), 1130–1144. https://doi.org/10.1111/1467-8624.00597

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114. https://doi.org/10.1016/j.tics.2008.11.008

Lu, L., Han, M., Zou, G., Zheng, L., & Gao, J.-H. (2022). Common and distinct neural representations of imagined and perceived speech. *Cerebral Cortex*, bhac519. https://doi.org/10.1093/cercor/bhac519

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718. https://doi.org/10.3758/PBR.16.4.711

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual Penetration of Visual Processing. *Psychological Science*, *21*(5), 682–691. https://doi.org/10.1177/0956797610366099

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.

Macbeth, A., Bruni, M. R., De La Cruz, B., Erens, J. A., Atagi, N., Robbins, M. L., Chiarello, C., & Montag, J. L. (2022). Using the Electronically Activated Recorder (EAR) to Capture the Day-to-Day Linguistic Experiences of Young Adults. *Collabra: Psychology*, *8*(1), 36310. https://doi.org/10.1525/collabra.36310

MacKay, D. G. (1981). The Problem of Rehearsal or Mental Practice. *Journal of Motor Behavior*, *13*(4), 274–285. https://doi.org/10.1080/00222895.1981.10735253

MacKay, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory imagery* (pp. 121–149). Erlbaum.

Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 436–448. https://doi.org/10.1037/0278-7393.21.2.436

MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: The neurobiology of sign language. *Trends in Cognitive Sciences*, *12*(11), 432–440. https://doi.org/10.1016/j.tics.2008.07.010

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591. https://doi.org/10.3758/BF03195849

Mandelbaum, E., Dunham, Y., Feiman, R., Firestone, C., Green, E. J., Harris, D., Kibbe, M. M., Kurdi, B., Mylopoulos, M., Shepherd, J., Wellwood, A., Porot, N., & Quilty-Dunn, J. (2022). Problems and Mysteries of the Many Languages of Thought. *Cognitive Science*, *46*(12). https://doi.org/10.1111/cogs.13225

Mani, N., & Plunkett, K. (2010). In the Infant's Mind's Ear: Evidence for Implicit Naming in 18-Month-Olds. *Psychological Science*, *21*(7), 908–913. https://doi.org/10.1177/0956797610373371

Marshall, P. H., & Cartwright, S. A. (1978). Failure to Replicate a Reported Implicit/Explicit Speech Equivalence. *Perceptual and Motor Skills*, *46*(3_suppl), 1197–1198. https://doi.org/10.2466/pms.1978.46.3c.1197

Marshall, P. H., & Cartwright, S. A. (1980). A final (?) note on implicit/explicit speech equivalence. *Bulletin of the Psychonomic Society*, *15*(6), 409–409. https://doi.org/10.3758/BF03334573

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. https://doi.org/10.1016/j.concog.2011.08.005

McCormick, A., Meijen, C., Anstiss, P. A., & Jones, H. S. (2019). Self-regulation in endurance sports: Theory, research, and practice. *International Review of Sport and Exercise Psychology*, *12*(1), 235–264. https://doi.org/10.1080/1750984X.2018.1469161

McCormick, A., Meijen, C., & Marcora, S. (2018). Effects of a motivational self-talk intervention for endurance athletes completing an ultramarathon. *The Sport Psychologist*, *32*(1), 42–50.

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *The Pavlovian Journal of Biological Science*, *24*(1), 19–26. https://doi.org/10.1007/BF02964527

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, *26*(1), 29–38. https://doi.org/10.1017/S0033291700033699

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.

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00285

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1518–1533. https://doi.org/10.1037/a0013355

Mischel, W., Cantor, N., & Feldman, S. (1996). Principles of self-regulation: The nature of willpower and self-control. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 329–360). Guilford.

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. https://doi.org/10.1016/j.actpsy.2003.12.004

Moffatt, J., Mitrenga, K. J., Alderson-Day, B., Moseley, P., & Fernyhough, C. (2020). Inner experience differs in rumination and distraction without a change in electromyographical correlates of inner speech. *PLOS ONE*, *15*(9), e0238920. https://doi.org/10.1371/journal.pone.0238920

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., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*(3), 376–382. https://doi.org/10.1002/acp.3404

Morin, A., Uttl, B., & Hamper, B. (2011). Self-Reported Frequency, Content, and Functions of Inner Speech. *Procedia - Social and Behavioral Sciences*, *30*, 1714–1718. https://doi.org/10.1016/j.sbspro.2011.10.331

Murray, D. J. (1965). The effect of white noise upon the recall of vocalized lists. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *19*(4), 333–345. https://doi.org/10.1037/h0082913

Nalborczyk, L. (2022). Re-analysing the Data From Moffatt et al. (2020): What Can We Learn From an Under-powered Absence of Difference? *Collabra: Psychology*, *8*(1), 33059. https://doi.org/10.1525/collabra.33059

Nalborczyk, L., Grandchamp, R., Koster, E. H. W., Perrone-Bertolotti, M., & Lœvenbruck, H. (2020). Can we decode phonetic features in inner speech using surface electromyography? *PLOS ONE*, *15*(5), e0233282. https://doi.org/10.1371/journal.pone.0233282

Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly Assessing the Relationship between Irrelevant Speech and Articulatory Suppression. *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1269–1278. https://doi.org/10.1080/02724980244000756

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. https://doi.org/10.1016/j.concog.2021.103102

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.

Newton, A. M., & de Villiers, J. G. (2007). Thinking While Talking: Adults Fail Nonverbal False-Belief Reasoning. *Psychological Science*, *18*(7), 574–579. https://doi.org/10.1111/j.1467-9280.2007.01942.x

Ngon, C., & Peperkamp, S. (2013). A deep look into the developing lexicon: Revelations from covert picture-naming. *Proceedings of the International Child Phonology Conference*, 29–30.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160. https://doi.org/10.3758/MC.38.8.1147

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.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pelizzon, L., Brandimonte, M. A., & Favretto, A. (1999). Imagery and Recognition: Dissociable Measures of Memory? *European Journal of Cognitive Psychology*, *11*(3), 429–443. https://doi.org/10.1080/713752323

Perona-Garcelán, S., Bellido-Zanin, G., Senín-Calderón, C., López-Jiménez, A. M., & Rodríguez-Testal, J. F. (2017). Spanish adaptation of the Varieties of Inner Speech Questionnaire (VISQ). Study of the relationship between inner speech, dissociation, and hallucination proneness. *Clínica y Salud*, *28*(2), 93–100. https://doi.org/10.1016/j.clysa.2017.02.001

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. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00122

Petitmengin, C., Remillieux, A., Cahour, B., & Carter-Thomas, S. (2013). A gap in Nisbett and Wilson's findings? A first-person access to our cognitive processes. *Consciousness and Cognition*, *22*(2), 654–669. https://doi.org/10.1016/j.concog.2013.02.004

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

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.

Postma, A., & Noordanus, C. (1996). Production and Detection of Speech Errors in Silent, Mouthed, Noise-Masked, and Normal Auditory Feedback Speech. *Language and Speech*, *39*(4), 375–392. https://doi.org/10.1177/002383099603900403

Ptacek, J. T., Smith, R. E., Espe, K., & Raffety, B. (1994). Limited correspondence between daily coping reports and restrospective coping recall. *Psychological Assessment*, *6*(1), 41–49. https://doi.org/10.1037/1040-3590.6.1.41

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Shankweiler, D. P., Katz, L., Fletcher, J. M., & others. (1996). Cerebral organization of component processes in reading. *Brain*, *119*(4), 1221–1238.

Pyers, J. E., & Senghas, A. (2009). Language Promotes False-Belief Understanding: Evidence From Learners of a New Sign Language. *Psychological Science*, *20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

Rao, K. V., & Baddeley, A. D. (2013). Raven's Matrices and Working Memory: A Dual-Task Approach. *Quarterly Journal of Experimental Psychology*, *66*(10), 1881–1887. https://doi.org/10.1080/17470218.2013.828314

Ratliff, K. R., & Newcombe, N. S. (2008). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology*, *56*(2), 142–163. https://doi.org/10.1016/j.cogpsych.2007.06.002

Ren, X., Wang, T., & Jarrold, C. (2016). Individual Differences in Frequency of Inner Speech: Differential Relations with Cognitive and Non-cognitive Factors. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01675

Resnik, P. (2018). Multilinguals' use of L1 and L2 inner speech. *International Journal of Bilingual Education and Bilingualism*, *24*(1), 72–90. https://doi.org/10.1080/13670050.2018.1445195

Richardson, J. T. E. (1987). Phonology and reading: The effects of articulatory suppression upon homophony and rhyme judgements. *Language and Cognitive Processes*, *2*(3–4), 229–244. https://doi.org/10.1080/01690968708406933

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

Saeki, E., Baddeley, A. D., Hitch, G. J., & Saito, S. (2013). Breaking a habit: A further role of the phonological loop in action control. *Memory & Cognition*, *41*(7), 1065–1078. https://doi.org/10.3758/s13421-013-0320-y

Saito, S. (1993). Phonological Similarity Effect is Abolished by a Silent Mouthing Task. *Perceptual and Motor Skills*, *76*(2), 427–431. https://doi.org/10.2466/pms.1993.76.2.427

Samuel, S., Durdevic, K., Legg, E. W., Lurz, R., & Clayton, N. S. (2019). Is Language Required to Represent Others' Mental States? Evidence From Beliefs and Other Representations. *Cognitive Science*, *43*(1), e12710. https://doi.org/10.1111/cogs.12710

Sandler, W. (2012). The Phonological Organization of Sign Languages: Sign Language Phonology. *Language and Linguistics Compass*, *6*(3), 162–182. https://doi.org/10.1002/lnc3.326

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, *23*(1), 117–141. https://doi.org/10.1348/026151004X21332

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review*, *14*(3), 527–533. https://doi.org/10.3758/BF03194102

Smith, S., Brown, H., Toman, J., & Googman, L. (1947). The Lack of Cerebral Effects of d-Tubocurarine. *Anesthesiology*, *8*(1), 1–14.

Snelgrove, R., & Havitz, M. E. (2010). Looking Back in Time: The Pitfalls and Potential of Retrospective Methods in Leisure Studies. *Leisure Sciences*, *32*(4), 337–351. https://doi.org/10.1080/01490400.2010.488199

Sokolov, A. N. (1968). *Inner speech and thought* (G. T. Onischenko, Trans.). Plenum Press.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Stets, A., Smith, S. L., & Helton, W. S. (2020). Dual-Task Interference Between Swimming and Verbal Memory. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(7), 1132–1140. https://doi.org/10.1177/0018720819871743

Talarico, M. K., Lynall, R. C., Mauntel, T. C., Weinhold, P. S., Padua, D. A., & Mihalik, J. P. (2017). Static and dynamic single leg postural control performance during dual-task paradigms. *Journal of Sports Sciences*, *35*(11), 1118–1124. https://doi.org/10.1080/02640414.2016.1211307

Theodorakis, Y., Weinberg, R., Natsis, P., Douma, I., & Kazakas, P. (2000). The effects of motivational versus instructional self-talk on improving motor performance. *The Sport Psychologist*, *14*(3), 253–271.

Thibodeaux, J., & Winsler, A. (2018). What do youth tennis athletes say to themselves? Observed and self-reported self-talk on the court. *Psychology of Sport and Exercise*, *38*, 126–136.

Thibodeaux, J., & Winsler, A. (2022). Stay Positive: Self-Talk in Relation to Motivational Climate, Goal Orientation, and Self-Talk Encouragement in Tennis. *Research Quarterly for Exercise and Sport*, *93*(4), 769–780. https://doi.org/10.1080/02701367.2021.1918323

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00314

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12. https://doi.org/10.1016/j.cortex.2016.01.002

Tod, D., Hardy, J., & Oliver, E. (2011). Effects of Self-Talk: A Systematic Review. *Journal of Sport and Exercise Psychology*, *33*(5), 666–687. https://doi.org/10.1123/jsep.33.5.666

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, *135*(2), 252–256. https://doi.org/10.1016/j.actpsy.2010.07.008

Uttl, B., Morin, A., & Hamper, B. (2011). Are Inner Speech Self-Report Questionnaires Reliable and Valid? *Procedia - Social and Behavioral Sciences*, *30*, 1719–1723. https://doi.org/10.1016/j.sbspro.2011.10.332

Van Raalte, J. L., Vincent, A., & Brewer, B. W. (2016). Self-talk: Review and sport-specific model. *Psychology of Sport and Exercise*, *22*, 139–148. https://doi.org/10.1016/j.psychsport.2015.08.004

Vicente, A., & Martínez-Manrique, F. (2016). The nature of unsymbolized thinking. *Philosophical Explorations*, *19*(2), 173–187. https://doi.org/10.1080/13869795.2016.1176234

Vygotsky, L. S. (1962). Thought and language. In *Thought and language* /. Massachusetts Institute of Technology. New York and London: John Wiley.

Vygotsky, L. S., & Luria, A. (1994). Tool and Symbol in Child Development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky Reader* (pp. 99–174). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

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. https://doi.org/10.1044/2017_JSLHR-S-16-0376

Wells, G. L., & Loftus, E. F. (2003). Eyewitness Memory for People and Events. In I. B. Weiner & A. M. Goldstein (Eds.), *Handbook of Psychology* (Vol. 11, pp. 149–160).

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the Time Course of Phonological Encoding. *Journal of Memory and Language*, *34*(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

Wilkinson, S., & Fernyhough, C. (2018). When Inner Speech Misleads. In *Inner Speech: New Voices* (Vol. 1, pp. 244–260). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0010

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.

Winsler, A., & Naglieri, J. (2003). Overt and Covert Verbal Problem-Solving Strategies: Developmental Trends in Use, Awareness, and Relations With Task Performance in Children Aged 5 to 17. *Child Development*, *74*(3), 659–678. https://doi.org/10.1111/1467-8624.00561

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition*, *35*(6), 1380–1398. https://doi.org/10.3758/BF03193609

Zourbanos, N., Hatzigeorgiadis, A., Bardas, D., & Theodorakis, Y. (2013). The Effects of Self-Talk on Dominant and Nondominant Arm Performance on a Handball Task in Primary Physical Education Students. *The Sport Psychologist*, *27*(2), 171–176. https://doi.org/10.1123/tsp.27.2.171

Article I

Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-022-02144-7

THEORETICAL/REVIEW

Verbal interference paradigms: A systematic review investigating the role of language in cognition

Johanne S. K. Nedergaard[1] · Mikkel Wallentin[1] · Gary Lupyan[2]

Accepted: 29 June 2022

Abstract

This paper presents a systematic review of the empirical literature that uses dual-task interference methods for investigating the on-line involvement of language in various cognitive tasks. In these studies, participants perform some primary task X putatively recruiting linguistic resources while also engaging in a secondary, concurrent task. If performance on the primary task decreases under interference, there is evidence for language involvement in the primary task. We assessed studies (N = 101) reporting at least one experiment with verbal interference and at least one control task (either primary or secondary). We excluded papers with an explicitly clinical, neurological, or developmental focus. The primary tasks identified include categorization, memory, mental arithmetic, motor control, reasoning (verbal and visuospatial), task switching, theory of mind, visual change, and visuospatial integration and wayfinding. Overall, the present review found that covert language is likely to play a facilitative role in memory and categorization when items to be remembered or categorized have readily available labels, when inner speech can act as a form of behavioral self-cuing (inhibitory control, task set reminders, verbal strategy), and when inner speech is plausibly useful as "workspace," for example, for mental arithmetic. There is less evidence for the role of covert language in cross-modal integration, reasoning relying on a high degree of visual detail or items low on nameability, and theory of mind. We discuss potential pitfalls and suggestions for streamlining and improving the methodology.

Keywords Working memory · Dual-task performance · Language/memory interactions

Introduction

Does language help us think and solve problems, and if so, how? What kinds of mental tasks depend most on the use of language? These classic questions, debated in philosophy and psychology for more than a century (Fodor, 1975; Müller, 1978; Sokolov, 1968; Vygotsky, 1962; Watson, 1913), have been increasingly tackled using various empirical and modelling methods (Baldo et al., 2005; Coetzee et al., 2019; Feinmann, 2020; Gilbert et al., 2006; Luo et al., 2021; Romano et al., 2018). One widely used method is verbal interference or articulatory suppression (Perry & Lupyan, 2013). In studies using this method, participants are asked to perform some task that may or may not require linguistic processing while at the same time performing a clearly linguistic task, such as repeating a word. If performance on the "primary" task is compromised by the verbal task more than by control non-verbal tasks, one can conclude that language in some form is likely to be recruited by the primary task. Specific studies using this paradigm (e.g., Hermer-Vazquez et al., 1999; Newton & de Villiers, 2007) become held up as evidence for the crucial role of language as a cognitive tool (Bermúdez, 2003; Carruthers, 2002; Clark, 1998; Gomila et al., 2012). But follow-up studies and (non)replications complicate the narrative, and the use of different types of verbal interference and different types of control conditions makes comparisons across areas difficult. Finding that verbal interference disrupts one task but not another is difficult to interpret if the types of verbal interference that were used are substantially different.

Given the complexity, diversity, and potential importance of this literature, it is valuable to systematically review the findings to date. There exist reviews that focus on some

✉ Johanne S. K. Nedergaard
nedergaard@cc.au.dk

[1] Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark

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

Published online: 22 August 2022

domains where language has been proposed to play a role: Gilhooly (2005) for the role of language in reasoning when using verbal materials, Kiesel et al. (2010) and Koch et al. (2018) for the role of language in task switching, DeStefano and LeFevre (2004) and Raghubar et al. (2010) for the role of language in mental arithmetic, Ratliff and Newcombe (2008) for the role of language in spatial reorientation, and Alderson-Day and Fernyhough (2015) for a narrative review of the cognitive functions of *inner speech* specifically. Still lacking, however, is a comprehensive review across areas. This paper aims to provide a one-stop shop for dual-task evidence of the role of language in cognition. Importantly, dual-task approaches are just one way to investigate the role of language in cognition. Other ways include introducing new verbal labels as an experimental manipulation, examining performance by speakers of different languages, or attempting to interfere with linguistic processes with TMS (transcranial magnetic stimulation) or tDCS (transcranial direct current stimulation). Verbal interference remains a common method for testing on-line (i.e., in-the-moment) involvement of language in cognition, and so it is the method we focus on here.

Objectives

Our primary goals were:

1. To provide a coherent overview to aid in understanding of what cognitive functions language may and may not be involved in.
2. To provide suggestions and recommendations for methodology used in future studies in order to make results from different experiments more comparable.
3. To provide theoretically motivated reasons for choosing one interference type over another.

Verbal interference and verbal working memory

Verbal interference was first used in studying working memory (Baddeley & Hitch, 1974; Murray, 1967; Peterson, 1969), specifically to test the hypothesis that there is a component of working memory dedicated to the processing and storage of verbal material (the phonological loop and the phonological store) (Baddeley, 2003). Articulatory suppression (a type of minimally demanding verbal interference in which participants repeat a syllable or short word out loud) was used to discover whether participants were using verbal rehearsal to maintain the memory trace of for example a series of letters. The assumption that the phonological loop or verbal working memory is a specialized part of working memory underlies most of the studies reviewed here. We exclude studies specifically investigating this claim, but all the included studies rely on different verbal tasks drawing on the same resources, and thus that we have such cognitive components dedicated to processing in a verbal format – an assumption that has been called into question (Baddeley & Larsen, 2007; Jones et al., 2004, 2007). Criticism of the assumption revolves around whether verbal working memory is verbal in an abstract sense or whether it simply involves low-level acoustic-articulatory processes. We omit discussion of this debate about the nature of "verbal" working memory because the logic of the dual-task design is valid regardless of the debate's outcome, even though it might be relevant when discussing how much of "language" different types of interference tasks plausibly interfere with.

In order to understand how verbal interference might work in more abstract cases, it is useful to first examine how it works in the most concrete, straightforward cases. Articulatory suppression has been used to investigate the so-called "phonological similarity effect" where serial recall performance is worse when the items to be remembered sound similar (Baddeley, 1966; Camos et al., 2013; Conrad, 1964; Conrad & Hull, 1964; Hintzman, 1967; Wickelgren, 1965a, b). The idea is that verbal working memory is divided into a phonological loop and a phonological store. Auditorily presented verbal material has direct access to the phonological store while verbal material presented visually (such as with written text) has to be converted in the phonological loop before it can enter the store. Thus, the phonological similarity effect should be different depending on presentation modality and the presence of articulatory suppression. See Fig. 1 for an illustration of an experiment testing the phonological similarity effect. Here, the hypothesis is that language is recruited to help store verbal material.

Because performing two tasks at the same time demands additional resources, performance under verbal interference must be compared to performance under an equivalently demanding but non-verbal dual-task condition. If verbal interference causes a more severe performance decrease than another distracting task equivalent in all other respects than the verbal, this would provide a causal argument for the presence of a linguistic component in the primary task. Articulatory suppression is often compared with the effect of foot tapping, another simple motor task that has been shown to be as attentionally demanding as articulatory suppression (Emerson & Miyake, 2003, Appendix A).

Outside the study of working memory components, verbal interference has been used to study, for example, task switching where the phonological loop is hypothesized to be recruited for self-cuing of whatever the relevant rule is, such as the common paradigm of switching between solving addition and subtraction problems. Here, verbal interference also impairs performance. In this specific case, the hypothesis would be that language is recruited to solve a task where it is necessary to maintain and update the relevant rule on each individual trial. This is similar to storing verbal

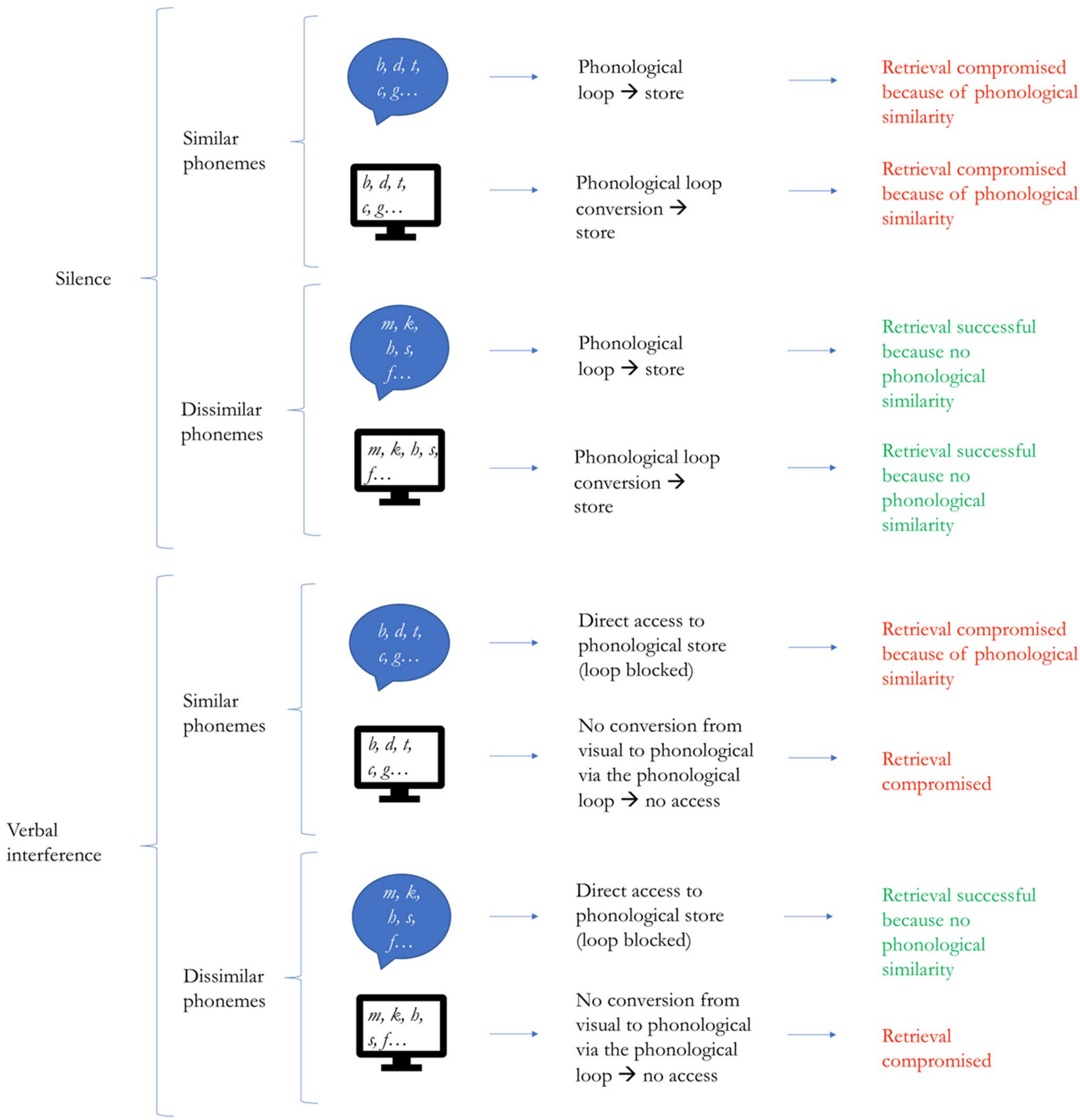


Fig. 1 A visualization of the mechanisms hypothesized to underlie the phonological similarity effect and how it differs depending on whether stimulus materials are presented verbally (speech bubble icon) or visually (screen icon)

materials in the phonological loop, except that instead of items to-be-remembered, the loop contains task instructions to-be-remembered. While covert language straightforwardly functions through verbal rehearsal in these examples, other studies have focused more on the structural and representational properties of language. These studies have used the dual-task interference methodology to test for example whether language aids cognition by providing the syntactic structure necessary for processing formal logic or by providing labelled categories that carve up otherwise continuous stimulus spaces. The precise mechanism of how repeating the word "December" (articulatory suppression) requires resources from the same cognitive component as recursive embedding and categorially labelled continua is less tangible than the precise mechanism of how articulatory suppression and task cuing might do the same. Similarly, many critics have pointed out the seeming paradox of how language can have "deep" effects on non-verbal cognition that are nevertheless disrupted by surface-level verbal interference (Dessalegn & Landau, 2008; Gleitman & Papafragou, 2005; Li et al., 2009; see Lupyan, 2012a, for a discussion).

Verbal interference across cognitive domains

The more abstract, structure- or representation-focused dual-task studies are of a very different flavor compared with purely rehearsal-focused studies that have delineated the precise mechanisms and sub-mechanisms very precisely. There is, for example, a long way from testing whether the phonological similarity effect persists under articulatory suppression as illustrated in Fig. 1 (see e.g., Jones et al., 2004) to testing whether something like false belief understanding relies on covert language (see e.g., Newton & de Villiers, 2007). The hypothesis here could for example be that theory-of-mind processing requires on-line access to sentential complements (e.g., 'She thinks [the apple is in the box]') but how verbal interference would block this access is less clear as it has not been shown that participants have to formulate the sentence 'she thinks the apple is in the box' explicitly in their minds to understand false belief on the fly. Thus, the easiest part of a study investigating the role of language in cognition with a dual-task experiment may be finding the effect – the more difficult part is explaining the precise mechanisms behind why this effect exists.

If there is one or several general roles that language plays in cognition, comparing the results of verbal interference across domains is one way of discovering what these might be. For example, most of the working memory-inspired studies included in the present review use very similar interference methods (word or syllable repetition) to test the role of covert language in task switching. By conducting slight variations on the primary task, these researchers thus zone in on whether covert language is recruited for task maintenance, task updating, task retrieval, etc. Once the precise effect is established, predictions are generated for other domains and we may test whether covert language also plays a role in for example task retrieval outside the addition/subtraction paradigm. Likewise, if we discover that verbal interference disrupts categorical perception of color, we should extend the paradigm to other types of categorical perception to ascertain whether covert language in general facilitates categorial perception. In the long term, it will of course also be necessary to integrate findings from other literatures apart from the dual-task interference literature (e.g., developmental evidence, evidence from brain lesions, evidence from noninvasive brain stimulation, etc.) As we proceed along this path, we can potentially map out domain-general functions of language for cognition, if such exist.

Review methodology

We followed PRISMA guidelines for selecting papers to include in this review (see Appendix B (OSM)). To be eligible, a paper needed to be peer-reviewed, and report at least one experiment with verbal interference and include at least one control task (either primary or secondary). Without such control tasks it is impossible to know whether the observed effects of verbal interference are purely due to the presence of a secondary task or whether they have something to do with language. We excluded studies in which the primary task being investigated was straightforwardly linguistic (e.g., lexical decision) because we were interested in the role of language in (putatively) non-verbal cognition. We also excluded papers with an explicitly clinical, neurological, or developmental focus. Although these studies are certainly valuable, including them would make it much more difficult to draw comparisons across areas, and so we leave their review for future work. We used the following search terms on PubMed and Google Scholar:

> *'articulatory suppression' OR 'dual-task paradigm' OR 'non-verbal control' OR 'verbal interference' NOT 'clinical' NOT 'developmental' NOT 'brain imaging'.*

To simplify the analysis of the findings, we divided the studies into clusters of primary task domains. If studies fitted into multiple clusters (e.g., if separate experiments within a study investigated different domains), the study is included in discussions of both clusters. For each study, the primary author collected the specific primary task(s), the specific interference task(s), the dependent variable(s), whether there was a selective effect of verbal interference, whether there was a difference between (levels of) the primary tasks, the number of participants in each experiment, and effect size(s) if reported. See Appendix A (OSM) for the full table including all the papers reviewed. The review was not registered, and a protocol was not prepared (aside from as detailed in the present section).

Results

Our literature search yielded 134 relevant papers, 33 of which were excluded (see criteria above), leaving 101 papers. We took great care to find as many of the relevant studies as possible, but as this literature is very fragmented and different subfields use different terminologies, we inevitably missed some. To the best of our knowledge, the present review represents an unbiased sample. We grouped the 101 relevant papers into 11 clusters based on the primary task: categorization (simple and complex), memory, mental arithmetic, motor control, reasoning (verbal and non-verbal materials), task switching, theory of mind, visual change, and visuospatial integration and wayfinding. In the following sections, we discuss the findings of the systematic review in terms of both the types of interference tasks used and the cognitive functions investigated.

Types of interference tasks

The several different types of interference tasks present their own challenges. It is sometimes unclear whether an effect is simply due to irrelevant aspects of the interference tasks, and it is thus necessary to include them in our discussions and analyses. Aside from syllable or word repetition (n = 61), the main types of verbal interference used are verbal short-term memory tasks (n = 22), verbal shadowing (n = 13), and verbal judgment tasks (n = 6). Each of these types is discussed below.

Syllable/word repetition

Syllable or word repetition is by far the most common type of verbal interference used in the literature reviewed here, found in 61 of the 101 studies. This kind of verbal interference is often referred to as "articulatory suppression" because it suppresses normal function of articulatory organs. Syllable or word repetition were the only types of verbal interference found to be used to disrupt the role of covert language in task switching (Baddeley et al., 2001; Brown & Marsden, 1991; Liefooghe et al., 2005; Weywadt & Butler, 2013). For example, in Emerson and Miyake (2003) participants were asked to complete lists of alternating arithmetic problems while engaging in either repetition of the phrase "a-b-c" once every 750 ms or tap their foot once every 750 ms. The comparison interference task is either foot tapping, simple finger tapping, or pattern finger tapping. In experiments with more visually detailed primary tasks than the alternating lists paradigm, syllable repetition tends to be compared with both simple tapping and pattern tapping. Although there are also different ways of using this kind of articulatory suppression, the ways are plausibly comparable (i.e., there is no a priori reason to believe that repeating "the" twice per second would be different from repeating another short, well-learned word at the same rate). One study investigated whether the semantic content of the words being repeated mattered for a navigational working memory task (Piccardi et al., 2020). The experimenters asked participants to repeat nonsense syllables, egocentric spatial words, or non-egocentric spatial words, and this study found no difference between the different classes of words being repeated.

Verbal memory

Twenty-two studies reviewed here used a memory-based concurrent task (Annett & Leslie, 1996; Cheetham et al., 2012; Clearman et al., 2017; Croijmans et al., 2021; Frank et al., 2012; Gilbert et al., 2006, 2008; He et al., 2019; Hegarty et al., 2000; Imbo & LeFevre, 2010; Kranjec et al., 2014; Liu et al., 2008; Lupyan, 2009; Maddox et al., 2004; Newell et al., 2010; Robert & LeFevre, 2013; Samuel et al., 2019; Trbovich & LeFevre, 2003; Vogel et al., 2001; Winawer et al., 2007; Witzel & Gegenfurtner, 2011; Zeithamova & Maddox, 2007). In memory-based concurrent tasks, participants are asked to engage in covert rehearsal of verbal and non-verbal materials during the primary task with a subsequent memory test. For example, Lupyan (2009) investigated thematic or perceptual odd-one-out judgment with word or picture stimuli as the primary tasks and verbal and visuospatial memory as the secondary interference tasks. The interference tasks were either a nine-digit verbal rehearsal with a four-alternative forced choice test after each trial or a nine-dot spatial rehearsal with a four-alternative forced choice test after each trial. Another frequent version of this memory-based verbal interference task is N-back matching, where words are presented sequentially and participants have to press a button if a word matches the one immediately preceding it (Gilbert et al., 2006, 2008; Kranjec et al., 2014; Liu et al., 2008). One issue with using memory tasks as interference is that it is difficult to separate the different stages of memory encoding. If there are interference effects, it is difficult to see whether this happens at the encoding, maintenance, or retrieval stages. It could be that participants simply encode and store the to-be-remembered material outside working memory (e.g., in long-term memory) at the beginning of a trial, especially when trials last more than a few seconds. This enables them to devote all of their verbal resources to the primary task until they have to retrieve the to-be-remembered material again after the trial.

Verbal shadowing

In verbal shadowing, participants are asked to "shadow" continuous speech – i.e. repeat as quickly as possible without breaks – while simultaneously performing a primary task. Compared to syllable repetition, verbal shadowing has been used in a wider range of experiments. It was for example used in three of the four theory-of-mind experiments reviewed here (Dungan & Saxe, 2012; Forgeot d'Arc & Ramus, 2011; Newton & de Villiers, 2007), one of the memory studies (Perkins & McLaughlin Cook, 1990), one study on motion events (Feinmann, 2020), one study on categorization (Simons, 1996), one study on number representation (Frank et al., 2012), and six out of ten of the studies on visuospatial integration and wayfinding (Bek et al., 2009, 2013; Hermer-Vazquez et al., 1999; Hupbach et al., 2007; Ratliff & Newcombe, 2005, 2008). For example, in Hermer-Vazquez et al. (1999), participants were asked to continuously shadow a tape recording of articles from a political newspaper. As a comparison interference task, Hermer-Vazquez et al. used a rhythm shadowing task where participants were asked to shadow-clap a sequence of clapped rhythm in 4/4 time that occurred at a rate of about 90 beats/min with a new rhythm played every eight beats. Rhythm shadowing is also used as the non-verbal interference task in the other studies using verbal shadowing.

The main difference between syllable repetition as discussed above and verbal shadowing is that verbal shadowing is arguably more demanding – to shadow successfully, you have to both perceive input and produce output at the same time. It is also less predictable and does not rely on overlearned sequences. Thus, the two verbal interference methods are not strictly comparable as verbal shadowing may target more aspects of natural language than simply the phonological loop.

Judgment tasks

Finally, six studies used judgment tasks as verbal interference, a more varied class of tasks that differ in their demands on response inhibition and comparisons between a presented stimulus and one (or several) held in memory. For example, Sims and Hegarty (1997) investigated "mental animation" (inferring the motions of mechanical systems) while having participants judge whether a specific letter was present in a list of six letters or not (putatively verbal interference) or decide if two patterns of four dots on a 4 × 4 grid were the same or different (a visuospatial interference condition). Hund (2016) and Meilinger et al. (2008) used similar interference tasks while examining wayfinding as the primary task. Here, the verbal interference task was word/non-word judgment. For the visual interference task, participants had to judge whether the two hands of the clock would be in the same half of the clock face or different halves of the clock face (dividing the clock face into an upper and a lower half) given a specific time of day (e.g., "6 o'clock"). Meilinger et al. (2008) also had a spatial interference task where participants were asked to judge from which direction a sound was coming. Pilling et al. (2003) used relative size discrimination and rhyme judgment. A special subclass of verbal judgment task is the Stroop task, where participants are presented with color words written with colored letters and have to respond based on the color of the letters and not the color name of the word. This type of judgment task was used in two studies, both testing motor control (Biese et al., 2019; Talarico et al., 2017).

Interim discussion of interference tasks

We found four main types of verbal interference tasks: syllable repetition, verbal memory, verbal shadowing, and judgment tasks. The review of the different tasks raised a few issues. First, it was not always clear to us which task was secondary and which was primary. Second, it is often difficult to assess performance on the interference task. Third, the verbal and non-verbal interference tasks do not always live up to the dual constraints of being (a) equally demanding and (b) different in only the presence or absence of "verbality" (Perry & Lupyan, 2013). We address these issues here.

In several studies we reviewed, it was unclear which was the "primary" task and which was the "secondary". Usually, researchers are interested in investigating the role of covert language in a specific cognitive component which they term the primary task (e.g., memory for facial expressions) and use a secondary task (e.g., rhyme judgments) to interfere with the primary task. Many times, however, the distinction between primary and secondary task is merely a question of terms. Trying to memorize facial expressions might interfere with rhyme judgments, but making rhyme judgments might also interfere with trying to memorize facial expressions. It is necessary therefore to measure a potential trade-off effect where participants may devote all their resources to the secondary task instead of the primary task – if there is a trade-off effect, performance on the primary task and performance on the secondary task should be negatively correlated. Unfortunately, this is very rarely reported and often cannot be assessed because performance on the secondary task is generally not measured. This is, for example, the case with syllable repetition and verbal shadowing where the experimenters do not objectively assess performance, often simply writing something to the effect of: 'The experimenters monitored that participants repeatedly uttered the word 'the' at 2 Hz.' Without having some form of performance measure on the secondary task, we have no way of knowing how engaged participants are in the task, and whether the engagement fluctuates according to the demands of the primary task, for example, participants may strategically pause shadowing or verbal rehearsal when faced with a difficult trial on the primary task.

The third issue relates to how comparable the verbal and non-verbal interference tasks are. Ideally, the two tasks should be simultaneously equally difficult and attentionally demanding and differ only in their involvement of language. This is difficult to operationalize and has not always been done (or done well). Hermer-Vazquez et al. (1999), for example, ascertained that their verbal shadowing and rhythm shadowing tasks were equally demanding by assessing participants' performance on a visual search task and finding that the two interference tasks had comparable detrimental effects. The conclusion that the two tasks are equally demanding in this case relies on the assumption that a visual search task would demand equal resources from verbal and visuospatial working memory, which is debatable. Relatedly with studies using syllable repetition, there has been some debate on whether the foot tapping task is an appropriate equivalent interference task in terms of demand. Proponents argue that it *is* equivalent because it is a simple motor task like repeating a word and should be as automatic and undemanding of the "central executive," the only difference between syllable repetition and foot tapping then being that syllable repetition involves articulatory organs (e.g., Emerson & Miyake, 2003, Appendix A).

In the discussions of the primary tasks investigated below, it is important to keep these interference task issues in mind. It may be the case that the presence or absence of verbal interference effects are not caused by the involvement or lack thereof of covert language but rather caused by incomparability of verbal and non-verbal interference tasks, hidden trade-off effects, or interference tasks that are not appropriate to the primary task investigated.

Effects of verbal interference on different cognitive tasks

We first describe the key studies from each family of primary functions we investigated and summarize the overall findings. The broad categories of primary functions investigated (ordered by how many studies each category contains) are: reasoning (verbal and non-verbal materials), memory, task switching, categorization (simple and complex), visuospatial integration and wayfinding, mental arithmetic, visual change, theory of mind, and motor control. See Appendix A (OSM) for a listing of the individual studies.

Reasoning

We identified 20 studies investigating reasoning. These can be divided into those using verbal materials (which encompasses studies that investigate formal logical problem-solving presented in a verbal format) and those using non-verbal materials (e.g., matrix reasoning, visual recursion, Tower of London).

Using verbal materials Eight studies investigated the role of covert language in reasoning using verbal materials (Evans & Brooks, 1981; Farmer et al., 1986; Gilhooly et al., 1993, 1999, 2002; Klauer, 1997; Meiser et al., 2001; Toms et al., 1993), which include propositional reasoning, conditional reasoning, and syllogistic reasoning. Here, covert language is hypothesized to help through providing a representational structure that facilitates reasoning with premises, conclusions, conditionals, assumptions, etc. Problems are presented in a verbal format and participants usually have to respond by saying whether the conclusion is valid or invalid.

Evans and Brooks (1981) tested participants on conditional reasoning and found that their rate of accepting invalid inferences was not affected by either simple, overlearned articulatory suppression (repeating the digits 1–6 in order) or articulatory suppression with a memory load (repeating the digits 1–6 in a random order specified by the experimenter). Somewhat surprisingly, response times were actually faster during articulatory suppression (this pattern is frequently seen; we comment on it in the Discussion). Testing both true/false judgments of declarative sentences about the order of two presented letters and mental rotation judgments, Farmer et al. (1986) found that digit repetition selectively impaired reasoning while spatial tapping selectively impaired the mental rotation judgments. In contrast with Evans and Brooks (1981), Toms et al. (1993) investigated conditional reasoning and found that articulatory suppression instantiated by repeating a simple overlearned sequence did not impair reasoning judgments, but that articulatory suppression with a memory load did. Specifically, the memory-load condition made participants less likely to accept valid modus tollens inferences (if p then q ➔ not q then not p). As Toms et al. (1993) themselves point out, there were some methodological differences between the two studies – most importantly, the study by Evans and Brooks used a between-subjects design, which could mean that it was not sufficiently sensitive to separate interference effects from individual differences in reasoning abilities.

Generally, the studies found a specific disruptive effect of random number generation but not of concurrent repetition of an overlearned sequence of digits. The latter was sometimes also disruptive – although the pattern is far from clear – but never more so than visuospatial concurrent tasks when these were included. Articulatory suppression seemed to be more disruptive when premises were presented sequentially than when they were presented simultaneously (see Gilhooly et al., 1993, 2002, respectively). Especially the finding that dual-task interference is observed with trained/skilled participants but not with untrained/low-skilled participants is relevant for the present review as an illustration of the idea that reliance on a verbal strategy in reasoning might depend on skill-level.

Using non-verbal materials Reasoning using non-verbal materials encompasses 12 studies, three of which included the Tower of London task as the primary task (Cheetham et al., 2012; Phillips et al., 1999; Wallace et al., 2017), two tested the Wisconsin Card Sorting Task (Baldo et al., 2005; Dunbar & Sussman, 1995), two tested a Visual Errands Test (Law et al., 2006, 2013), one tested paper folding, card rotations, and picture matching (Hegarty et al., 2000), one tested visual recursion (Martins et al., 2015), one tested the Hidden

Figures Test (Miyake et al., 2001), one tested Raven's Progressive Matrices (Rao & Baddeley, 2013), and one tested analogical mapping (Waltz et al., 2000). Generally, in these cases, language is hypothesized to be involved as a problem-solving tool where participants discuss with themselves or simulate potential solutions to the problems internally. It is also sometimes the case that covert language is hypothesized to help by providing a label for the rule when this has to be discovered (e.g., in the Wisconsin Card Sorting Task, in the Martins et al. visual recursion study, or in Raven's Progressive Matrices).

The Tower of London task requires participants to move a stack of discs from one peg to another while preserving a specific order (e.g., a smaller disc can never be under a larger disc). Of the three studies investigating the Tower of London task, only Wallace et al. (2017) found a specific effect of articulatory suppression with participants making more excess move in this condition. Cheetham et al. (2012) used memory-based interference tasks and found that only performance on the secondary tasks was affected – and not performance on the Tower of London task. Visuospatial memory was significantly worse when performed concurrently with the Tower of London task. Notably, Phillips et al. (1999) found that articulatory suppression had a *positive* effect on both completion time and error rate.

The Wisconsin Card Sorting Task requires participants to sort cards according to rules that they have to discover through trial-and-error and which change frequently. Dunbar and Sussman (1995) found a specific effect of articulatory suppression on perseverative errors (when participants persevere with sorting according to a rule that has changed) compared with tapping but no interference on number of categories achieved or non-perseverative errors. In contrast, Baldo et al. (2005) found that both articulatory suppression and foot tapping were associated with more perseverative and non-perseverative errors, but these two interference conditions were importantly not statistically different from each other. This means that we cannot say if the impairment was due to dual-task demands or specifically due to verbal demands.

The Visual Errands Test did not appear to be affected by verbal interference. In this kind of study, participants must complete a list of errands in a virtual environment while taking care not to break some rules. Thus, this task is more about planning and multitasking than about visuospatial orientation. In both studies (Law et al., 2006, 2013), the interference tasks hypothesized to involve the Central Executive (random month generation, tone localization) had larger negative impact than articulatory suppression. There was no specific effect of verbal interference on number of errands completed, number of errors/rule breaks, or time.

The remaining four studies in this section investigated the Hidden Figures Test (Miyake et al., 2001), Raven's Progressive Matrices (Rao & Baddeley, 2013), visual recursion (Martins et al., 2015), and analogical mapping (Waltz et al., 2000) respectively. The Hidden Figures Test is a visuospatial problem-solving test requiring participants to identify which of five simple figures is hidden inside a more complex figure. In Raven's Progressive Matrices, participants are presented with a set of patterns organized according to a specific rule, and need to figure out which of several patterns best completes a 3 × 3 matrix. In Martins et al. (2015)'s study, participants were asked to judge whether some visual patterns could be generated by recursive rules from other visual patterns. In the analogical mapping task investigated by Waltz et al. (2000), participants have to map visual scenes onto each other by their relational properties instead of their surface properties. None of these four studies showed a specific negative effect of verbal interference.

Taken together, verbal interference does not obviously disrupt visuospatial problem-solving of the kind tested in these studies. Only two of the 12 studies – Dunbar and Sussman (1995) and Wallace et al. (2017) – found a specific disruptive effect of verbal interference. Interestingly, in both Dunbar and Sussman (1995) and Wallace et al. (2017), verbal interference was associated with less inhibitory control, i.e., making more excess moves or continuing with perseverative errors. This may indicate that covert language is recruited for inhibitory control.

Memory

We found 17 studies that investigated memory under different interference conditions (Annett & Leslie, 1996; Brandimonte et al., 1992a, b; Croijmans et al., 2021; Gaillard et al., 2012; Gimenes et al., 2016; Henson et al., 2003; Hitch et al., 1995; Mitsuhashi et al., 2018; Nakabayashi & Burton, 2008; Pelizzon et al., 1999; Perkins & McLaughlin Cook, 1990; Souza & Skóra, 2017; Vandierendonck et al., 2004; Vogel et al., 2001; Walker & Cuthbert, 1998; Wickham & Swift, 2006). Covert language is hypothesized to aid memory in different ways, for example by providing a more abstract code for the item to be remembered in addition to the representation in the relevant sensory modality (Paivio, 1991). This is known as *dual coding theory* and posits that a memory trace is stronger if it is captured by both perceptual experience and verbal experience. Alternatively, covert language could aid memory by providing a medium for continuous rehearsal of the items to be remembered. Of course, these two hypotheses are not mutually exclusive as covert language could potentially aid memory both by encoding and by rehearsal.

Henson et al. (2003) did not find a specific detrimental effect of articulatory suppression on either a list probe task assessing memory for serial order of visually presented letters or on an item probe task assessing memory for single item presence or absence. The three other interference tasks were irrelevant sound presentation, simple finger tapping, and complex, syncopated finger tapping. There was some indication that irrelevant sound and articulatory suppression had a larger detrimental effect on the list probe task than on the item probe task, although this was likely due to a ceiling effect on the item probe task. Thus, the results from Henson et al. (2003) do not support a selective role of covert language in either memory for either serial order or individual items. On the other hand, Nakabayashi and Burton (2008) reported a specific detrimental effect of articulatory suppression on facial recognition memory. Articulatory suppression during encoding was associated with worse performance on recognition memory compared with both a verbalization condition (where participants were asked to describe the faces out loud) and a simple tapping condition. Interestingly, Experiment 4 of Nakabayashi and Burton (2008) showed some indication that encoding the faces verbally *after* visual presentation had a weak detrimental effect on recognition memory. This suggests that the benefits of verbal encoding of visual stimuli depend on timing – this is reminiscent of the verbal overshadowing effect (Schooler & Engstler-Schooler, 1990), which is the finding that (forced) verbal descriptions of visual stimuli make subsequent recognition memory worse. In fact, Wickham and Swift (2006) investigated the verbal overshadowing effect specifically and found that verbal interference during stimulus presentation made the detrimental effect of subsequent verbal (over)description disappear.

Investigating memory for gestures, Gimenes et al. (2016) found that a verbal strategy (training manipulation) for remembering gestures was better than a gestural strategy, and that verbal interference interfered with gesture reproduction accuracy regardless of strategy. In a similar study, Mitsuhashi et al. (2018) found a specific effect of verbal interference on the Luria Hand Test, which measures reproduction accuracy. Less conclusive evidence for the facilitative role of language in memory comes from Walker and Cuthbert (1998), who investigated memory for color-shape associations, only using articulatory suppression as an interference task – thus it is not possible in this case to tell if there was a specific effect or not. However, they found that articulatory suppression disrupted the nameability advantage associated with some of the stimuli, supporting the idea that linguistic labelling facilitates memory. Interestingly, Souza and Skóra (2017) also found that overtly labelling colors to be remembered facilitated reproduction accuracy but also made the memory representation more categorical – in contrast, concurrent syllable repetition had a detrimental effect on reproduction accuracy.

Four of the memory studies tested the effect of verbal interference on both recognition memory and mental transformations of images (Brandimonte et al., 1992a, b; Hitch et al., 1995; Pelizzon et al., 1999). These studies found that while verbal interference disrupted recognition memory, mental transformation of the images to be remembered was actually improved by verbal interference. Mental transformation in this case refers to subtracting elements from the images, rotating them, or combining them to produce other recognizable forms. In addition, both advantages and disadvantages (e.g., stemming from degree of nameability) associated with verbal labelling disappeared with verbal interference. The authors of these four studies interpret the findings to mean that we normally use verbal resources to name visual stimuli to be remembered, and that this helps us recognize the stimuli later. However, the stored representation in verbal format does not maintain all the details of the original visual stimuli, which is why manipulations that depend on visual details are easier under verbal interference. This interpretation fits well with the color memory study by Souza and Skóra (2017) discussed above.

In most memory studies, the material to be remembered is presented visually, and nameability effects are found. However, some studies have also investigated the olfactory modality and memory for odors. Olfactory memory has been argued to depend on both a verbal code (taking advantage of odor labels) and a visual code (encoding an odor as the image of an object that prototypically smells like that). In a study that tested memory for wine odors, Croijmans et al. (2021) found that while experts were better than novices at both recognition and free recall, verbal interference had no effect on either group. Of the two other olfactory memory studies, one also did not find that verbal interference negatively affected memory performance (Annett & Leslie, 1996) and one found that digit shadowing had a specific negative effect on recognition, but not free recall (Perkins & McLaughlin Cook, 1990). Thus, there is no firm support for the on-line role of covert language in olfactory memory.

In summary, encoding items to be remembered verbally can be both beneficial (e.g., nameability advantages) and detrimental (e.g., verbal overshadowing effect), depending on what is to be remembered. The studies discussed here appear to support the idea that covert language influences memory as both advantageous and disadvantageous effects associated with verbal encoding disappeared under verbal interference.

Task switching

The present review found 16 studies investigating the role of covert language in task switching (Baddeley et al., 2001; Brown & Marsden, 1991; Bryck & Mayr, 2005; Emerson & Miyake, 2003; Grange, 2013; Kirkham et al., 2012; Liefooghe et al., 2005; Miyake et al., 2004; Saeki, 2007; Saeki et al., 2006, 2013; Saeki & Saito, 2004a, b, 2009;

Tullett & Inzlicht, 2010; Weywadt & Butler, 2013). All these studies test participants' ability to switch between two tasks and measure switch cost on reaction time and error rate (i.e., how much slower are the responses when a task B trial immediately follows a task A trial compared to if it follows another task B trial). These tasks included adding and subtracting numbers (Baddeley et al., 2001; Emerson & Miyake, 2003; Saeki & Saito, 2004a), color or shape sorting tasks (Kirkham et al., 2012; Liefooghe et al., 2005; Miyake et al., 2004), numerical or physical size judgment tasks (Saeki, 2007; Saeki et al., 2006, 2013; Saeki & Saito, 2004b, 2009), a Stroop task (Brown & Marsden, 1991), arithmetic problems verification (Bryck & Mayr, 2005), detection of different visual shapes preceded by visual cues (Grange, 2013), switched and regular versions of a Go/No-go task (Tullett & Inzlicht, 2010), and voluntary switching between odd/even and high/low digit judgments (Weywadt & Butler, 2013). It is worth noting that it is difficult to say if these task-switching experiments investigate flexibility (as participants need to flexibly shift between task sets) or inhibition (as participants need to inhibit the responses that they would make according to the non-active task set), or indeed if these two processes are two sides of the same coin.

As is evident from the above list, there are several different types of switch tasks represented in this primary task category – however, they all have in common that participants are asked to switch between responding to the same stimuli according to the rules of two different task sets. Usually, the studies also compare conditions where the relevant rule is somehow cued (e.g., displaying a '+' when the task is to add and a '−' when the task is to subtract) to conditions where the relevant rule is not cued or cued in a different way (e.g., endogenously vs. exogenously). Participants are hypothesized to retrieve and maintain the relevant rule or task set verbally. When the relevant rule is externally cued, articulatory suppression should have no effect if verbal rehearsal is under normal circumstances used as a sort of internal cue. Additionally, the studies also all use syllable repetition and foot or finger tapping as verbal and non-verbal interference tasks.

As an example of one of these task-switching studies, Baddeley et al. (2001) conducted seven experiments where they varied the types of interference task while participants completed either blocked or switched lists of numbers to be added or subtracted. The task on an individual trial either required the participant to remember the rule (endogenous condition) or included the rule as indicated by a plus or a minus sign (exogenous condition). Performance on switched trial lists was slower than on blocked trial lists – the experimenters measured the cumulative reaction time on a list where the participants had to alternate between adding and subtracting 1 and a list where they always had to either add or subtract 1. There were two different interference tasks as well: articulatory suppression (reciting days of the week or months of the year) and task taxing the central executive *and* verbal working memory (alternating day of the week and month of the year; Monday – January – Tuesday – February etc.). The executive task was associated with slower performance on both switched and blocked trials while articulatory suppression only appeared to slow performance on switched trials. Further, reaction times were slower with verbal interference on endogenously versus on exogenously cued trials. This difference between reaction times presumably indicates the cost associated with maintaining and drawing on a mental representation of the task (adding or subtracting).

Overall, the pattern of results from these 16 studies supports the idea that covert language is used to retrieve and maintain the task-relevant rule. Articulatory suppression seems to disrupt task switching when task cues are not present in the stimuli (Emerson & Miyake, 2003), suggesting that verbal rehearsal is needed to "remind" the participant of the task at hand.

Categorization

Sixteen studies investigated the role of language in categorization (Gilbert et al., 2006, 2008; He et al., 2019; Liu et al., 2008; Lupyan, 2009; Maddox et al., 2004; Minda et al., 2008; Newell et al., 2010; Pilling et al., 2003; Roberson & Davidoff, 2000; Winawer et al., 2007; Witzel & Gegenfurtner, 2011; Zeithamova & Maddox, 2007). In categorization studies, covert language is hypothesized to aid cognition by providing labels to carve up continuous perceptual space, for example, the color spectrum (Lupyan, 2012a). In studies that investigate novel category learning, covert language is supposedly recruited for learning discrimination patterns that are rule-based and easily verbalizable. In contrast, discrimination patterns that rely on more high-dimensional patterns are hypothesized to be learned in a more procedural way (see e.g., Maddox & Ashby, 2004). There are important differences between studies where participants need to categorize along some criterion (e.g., that does not belong based on size) and odd-one-out/perceptual matching studies. These tasks vary a great deal in how much you need to know to perform well, for example, detecting a visual difference versus using semantic knowledge or learned rules to solve a given categorization problem. Therefore, we divide this section into "simple categorization" and "complex categorization". The first section includes studies investigating perceptual discrimination and matching within and between known categories. The second section includes studies that involve learning novel categories and forming ad hoc categories involving, for example, focusing on one dimension while abstracting over other dimensions.

Simple categorization These studies investigate the use of already existing categories for detection of differences (e.g., between different colors). Most of them focus on color categories, although the categorization of facial expressions, spatial relations, and animals have also been investigated. In the color classification studies, participants are presented with a color and asked to classify it or presented with a selection of colors and asked to find the odd one out. In Gilbert et al. (2006), for example, participants were presented with a circle of colored squares where all except one were the same color. Participants then had to respond indicating which half of the circle the odd colored square was in. The color of the odd square was either in the same color category as the remaining squares (e.g., a different shade of green) or in a different color category (e.g., blue among greens). This study found that there was a cross-category advantage in the right visual field, possibly related to verbal labels, but that this advantage disappeared under verbal interference. A later study, however, attempted to replicate the Gilbert et al. (2006) findings but found that if the colors were more carefully controlled, the effect of visual field disappeared and did not differ depending on the presence or absence of verbal interference (Witzel & Gegenfurtner, 2011). Other studies without verbal interference have successfully replicated the visual field effect (Zhong et al., 2015; Zhou et al., 2010). In a study testing Russian- and English-speaking participants, Winawer et al. (2007) found the two groups differed when they were asked to discriminate shades of blue that were either within-category or across-category for the Russian speakers (Russian "blue" is divided into two separate terms, "goluboy" meaning lighter blues and "siniy" meaning darker blues). There was a category advantage for Russian speakers but not for English speakers. The Russian category advantage disappeared with verbal interference. A parallel effect was found by He et al. (2019), who tested Chinese and Mongolian speakers (the latter have different color words for light blue and dark blue, the former do not). Extending the category effects found in color discrimination, Gilbert et al. (2008) investigated categorization of dog and cat silhouettes and found that the language-based categorization effect was stronger in the right visual field than in the left, and that this category effect was attenuated by verbal interference.

Kranjec et al. (2014) tested categorical and coordinate spatial relation tasks and found that a one-back word-matching task had a larger disruptive effect than a one-back pattern-matching task. In these spatial relations tasks, participants were asked to make same/different judgments of dot-cross configurations that differed in how verbalizable the differences were. Counter to the author's prediction, there was no difference between the effect of verbal interference on trials with easier-to-name versus harder-to-name spatial categories. Two other studies investigating categorical and coordinate spatial relation tasks did not find specific effects of verbal interference (Dent, 2009; van der Ham & Borst, 2011). These two both used syllable repetition as the interference task, although only one (van der Ham & Borst, 2011) also included a non-verbal interference task (finger tapping).

Investigating categorical perception of both color and faces, Roberson and Davidoff (2000) found a selective interference effect of a verbal concurrent task. With the verbal concurrent task, the increased accuracy usually associated with cross-category judgments relative to within-category judgments had disappeared. The authors interpret this as indicating that the advantages associated with categorical perception and memory of faces and colors derive from verbal encoding and storage. In an attempt to replicate Roberson and Davidoff's (2000) experiment, Pilling et al. (2003) found that if the type of interference task was unpredictable, the category advantage survived verbal interference. The authors suggest that unpredictability of interference task condition may have discouraged the use of a verbal strategy. In another study that similarly calls into question the role of on-line language in categorical perception of color, Liu et al. (2008) found that the cross-category boundary advantage survived verbal interference. Although these studies show somewhat conflicting results, they indicate some tentative support overall for the idea that linguistic labels facilitate the speed and accuracy with which we make discrimination and detection judgments.

Complex categorization In one group of studies, participants are asked to learn novel categories where the category structure is either rule-based and easily verbalizable (e.g., 'red things are in category A, blue things are in category B') or where the category structure relies on information-integration (where at least two differently expressed dimensions need to be combined) and is not easily verbalizable. Support for this distinction comes for example from Maddox et al. (2004), who found that a four-digit memory task disrupted the learning of rule-based category structures but not information-integration category structures. Similarly, Minda, Desroches, and Church (2008) found that adults under verbal interference displayed a category-learning pattern similar to that of children in that they found disjunctive rules harder to learn ('red and small OR blue and large things are in category A, blue and small things OR red and large things are in category B'). Zeithamova and Maddox (2007) found that both a visual and a verbal concurrent memory task disrupted rule-based category learning but not information-integration category learning. In interpreting the results of these studies, it is important to take into account that Newell et al. (2010) found that the dissociation between information-integration and rule-based categorization disappeared when only participants who actually learned the rule were included in the analysis.

In a study investigating complex processing of already learned category structures, Lupyan (2009) investigated effects of verbal and visuospatial interference on participants' ability to appreciate different kinds of similarities among pictures of familiar objects (or words denoting those objects). Participants were shown three pictures or words and asked to choose the object/word that was most different from the two based on its real-world color, size, or thematic/function relationship. The study was based on prior work showing that individuals with aphasia were selectively impaired when asked to isolate specific perceptual dimensions such as color or size, but were similar to controls when asked to group on more thematic or functional criteria (Cohen et al., 1980; Davidoff & Roberson, 2004; De Renzi & Spinnler, 1967; see Vignolo, 1999, for review). Lupyan sought to determine whether a similar dissociation could be observed in non-aphasia participants whose language was interfered with during the task, and found that verbal interference selectively affected color and size trials for both picture and word stimuli.

Visuospatial integration and wayfinding

Twelve studies investigated the role of covert language in visuospatial integration and wayfinding (Bek et al., 2009, 2013; Caffò et al., 2011; Garden et al., 2002; Hermer-Vazquez et al., 1999; Hund, 2016; Hupbach et al., 2007; Labate et al., 2014; Meilinger et al., 2008; Piccardi et al., 2020; Ratliff & Newcombe, 2005, 2008). In these studies, covert language is supposed to help by providing a common medium for the integration of information from different sensory modalities as well as different types of information from the same sensory modality (e.g., shape and color).

Hermer-Vazquez et al. (1999) is one of the most famous studies in this field and widely cited in philosophy of cognitive science as evidence for the role of language in cognition (Carruthers, 2002; Clark, 1998; Gomila et al., 2012). In the original study, participants were placed in a rectangular room and saw something being hidden in one of the corners of the room. They were then blindfolded and spun around until they were thoroughly disoriented. The dependent variable in this kind of study is participants' search behavior – which corner do they search in? How do they reorient themselves? Originally, Hermer-Vazquez et al. (1999) found that participants engaged in verbal shadowing were unable to combine geometric and color features of the room to find the right corner (i.e., using both the fact that two walls were shorter than the others and the fact that one end wall was painted a different color).

Six of the remaining studies reviewed include attempts to replicate and extend these findings, unsuccessfully in all cases. To test whether the size of the room mattered, both Hupbach et al. (2007) and Ratliff and Newcombe (2008; Experiment 3) used a bigger room than Hermer-Vazquez et al. (1999), and found that only a spatial interference task impaired reorientation performance. Bek et al. (2009) compared prose shadowing and syllable shadowing and found that neither reduced performance to chance levels as in Hermer-Vazquez et al. (1999). Testing the effect of the specific instructions given to participants, Ratliff and Newcombe (2005) tested the difference between implicit and explicit directions and found no specific effect of verbal interference. Similarly, Bek et al. (2013) found that prose and syllable shadowing both only disrupted reorientation performance when instructions were vague and non-specific like in Hermer-Vazquez et al. (1999). There was no difference between the two shadowing types. Further variations of the original paradigm include a study by Caffò et al. (2011) that tested a virtual version of the reorientation task with syllable repetition as the verbal interference task and spatial tapping as the spatial interference task. Performance during both interference tasks was worse than the control condition, but spatial interference was significantly worse than verbal interference. There is a risk, however, that this was a motor artifact – participants had to perform spatial tapping with the left hand and navigate the virtual environment with a joystick with the right hand.

The remaining five experiments in this category investigated wayfinding in various more complex ways. Labate et al. (2014) examined learning of maps including landmarks and routes through navigation in a real environment and found that a spatial tapping task was worse for performance than a syllable repetition task. Comparable results were found by Meilinger et al. (2008) and Hund (2016), who investigated similar wayfinding tasks with similar interference tasks, namely word/non-word judgments as the verbal interference and clock hand judgments as the visual interference. Both studies found that the visuospatial interference tasks had a stronger detrimental effect on performance than the verbal interference tasks. Potentially shedding light on the different contributions of visuospatial and verbal working memory, Garden et al. (2002: Experiment 2) found that the degree to which participants were affected by verbal and visuospatial interference tasks in a real-world navigation problem depended on individual differences in spatial ability. Specifically, participants with high spatial ability were more affected by a concurrent spatial tapping task, and conversely participants with low spatial ability were more affected by a concurrent verbal interference task. Further testing the effect of many different kinds of interference tasks, Piccardi et al. (2020) investigated navigational working memory and found that only sound localization disrupted performance. The other interference tasks were stationary walking, stationary complex movements, nonsense syllable repetition, repetition of egocentric spatial words, and repetition of non-egocentric spatial words.

Despite early findings, the studies discussed in this section taken together do not provide strong support for the idea that covert language is recruited for visuospatial integration and wayfinding.

Mental arithmetic

Nine studies investigated cognitive processes related to mental arithmetic and exact number representation (Clearman et al., 2017; Frank et al., 2012; Imbo & LeFevre, 2010; Lee & Kang, 2002; Logie et al., 1994; Robert & LeFevre, 2013; Seitz & Schumann-Hengsteler, 2000, 2002; Trbovich & LeFevre, 2003). The phonological loop is hypothesized to help with mental arithmetic by keeping track of partial results needed for further computations (Ashcraft, 1995; Imbo et al., 2005). The studies often contrast arithmetic problems that require fact retrieval (usually small problems < 10) and problems that require carry operations. Most of the studies in this section found that verbal interference disrupts mental arithmetic across varying presentation formats (auditorily, visually, horizontally, vertically), problem size, and kind of mental arithmetic (addition, subtraction, multiplication). However, testing the effect of different distractors, Clearman et al. (2017) found that attending to the color and location of three dots for subsequent recall had a larger adverse effect on the speed of mental arithmetic than attending to words presented aurally for subsequent recall. Thus, there was no evidence of specific verbal involvement. Frank et al. (2012), on the other hand, found that both verbal shadowing and a memory task disrupted exact number representation for larger quantities. They conducted three experiments, only one of which included a control interference task – a comparison between memory for a sequence of consonants and a sequence of dot locations on a grid. Taken together, these studies seem to indicate that covert language resources are recruited for mental arithmetic problems that are most effectively solved using a verbal code – this includes problems featuring carry and borrow operations, problems presented horizontally (contrasting with vertically presented problems that appear to invite visual strategies), and problems presented auditorily.

Visual change

The six studies in this category include those investigating visual change detection (Hollingworth, 2003; Sense et al., 2017; Simons, 1996), mental animation (Sims & Hegarty, 1997), similarity ratings of motion events (Feinmann, 2020), and visuospatial construction and memory (Bek et al., 2009: Experiment 1). Bek et al. (2009) found a specific detrimental effect of verbal interference, but this effect was limited to one of their tasks. They used a block design task in which participants were asked to construct two-dimensional designs of red and white blocks, and a complex figure task in which participants were asked to copy a figure and draw it again from memory after a delay. Verbal shadowing only interfered with the complex figure task and only if participants were shadowing during the encoding stage and not the retrieval stage. The authors argue that the reason verbal shadowing interfered with the complex figure task and not the block design was that the complex figure task contained nameable elements. Nameability was also an important factor in Simons (1996) where the advantage associated with change detection for common objects (hats, chairs, etc.) disappeared with verbal shadowing. Interestingly, Hollingworth (2003) compared detection of rotation change and token change and found that token change detection was in fact more accurate with verbal interference than in a control condition.

Theory of mind

Four studies have investigated the on-line role of covert language in theory of mind (Dungan & Saxe, 2012; Forgeot d'Arc & Ramus, 2011; Newton & de Villiers, 2007; Samuel et al., 2019). Theory of mind refers to the ability to attribute thoughts, beliefs, intentions, etc. to other humans, even when these are at odds with one's own worldview. The connection between language and theory of mind is a much debated topic with input from developmental psychology (Lohmann & Tomasello, 2003), evolutionary psychology (Dunbar, 1998; Malle, 2002), and neuroscience (Siegal & Varley, 2006), among others. One hypothesis for why language would aid theory of mind is that the syntactic structure of sentential complements is recruited for representing other people's mental states, for example, 'she thinks [that the apple is in the box]' (de Villiers, 2007; de Villiers & de Villiers, 2000; de Villiers & Pyers, 2002). Alternatively, the connection between theory of mind and language in development could be that hearing adults talk about mental states directs children's attention to unseen mental states as well as the abstract properties that superficially different situations have in common (Milligan et al., 2007).

Of the four studies reviewed here, only Newton and de Villiers (2007) found a specific effect of verbal interference on a theory-of-mind task where participants were asked to choose the correct ending for false belief videos. There was no effect of either verbal shadowing or rhythm shadowing (the comparison task) on true-belief videos. There are some issues with this experiment, however. For example, the authors did not include a control condition with no interference or attempt to equate the two interference tasks for difficulty. This latter point was rectified by Dungan and Saxe (2012), who found that when the verbal and non-verbal interference conditions were better equated for difficulty, there was no effect of verbal interference on false belief reasoning. Similarly, Forgeot d'Arc and Ramus (2011) compared belief judgment tasks and mechanistic judgment tasks, and found that verbal shadowing had an overall effect on performance but not specifically on belief attribution. They did not compare with another interference task. Testing the effect of a different type of verbal interference task, Samuel et al. (2019) compared performance on false belief and false-photograph trials with interference tasks that involved an eight-digit covert rehearsal with a memory test and a 4 × 4 grid pattern rehearsal with a memory test. This study did not find that the false belief task was specifically impaired by the verbal interference task. It is worth noting that the interference here was not during the encoding stage but instead between encoding and retrieval. Nevertheless, the results of these four studies seem to indicate that there is little evidence that covert language is involved in on-line theory-of-mind reasoning.

Motor control

We found two studies that investigated the role of covert language in motor control in some way: jump landing performance (Biese et al., 2019) and single leg postural control (Talarico et al., 2017). The reasoning behind why covert language would help with motor control stems from Vygotskian self-regulation, according to which we use our inner voice to control our own behavior (Vygotsky, 1962). Covert language focuses attention on motor control and can be used to cue specific subcomponent motor actions that facilitate the overall movement goal (e.g., jumping, serving, hitting, etc.). Both studies found that a verbal interference task had a specific disruptive effect, one on reaction time (Biese et al., 2019) and one on squatting speed and depth (Talarico et al., 2017). Both studies compared physical performance during a Stroop Color Word test versus on a Brooks Visuospatial task, but these two interference tasks are not necessarily equated in other respects than the verbal (see Judgment tasks section above). This lack of comparability is underscored by the fact that both the Stroop Color Word test and a Symbol Digit Modalities test (basically an association memory test) had adverse effects on jump landing performance in Biese et al.'s (2019) study. Thus, there is some doubt as to whether it was the verbal component of the Stroop task that caused the interference or just attentional demands – the Stroop task also is not "pure" verbal interference in that sense as it also puts demands on executive control (response inhibition).

Discussion

As the above review has illustrated, the literature investigating the role of covert language in cognition using dual-task methodologies is broad and varied. Nevertheless, it is possible to extract some general trends and tendencies. In the above sections, we provided an overview to aid in understanding what cognitive functions language may and may not be involved in. In the following, we will attempt to tie it all together. Additionally, we will provide suggestions and recommendations for methodology used in future studies – in order to make results from different experiments more comparable – and encourage theoretically motivated reasons for choosing one interference type over another.

Summary of the findings

As can be seen in Table 1 and Fig. 2, it seems to be the case that verbal interference has a specific disruptive effect on tasks involving simple categorization, mental arithmetic, memory, motor control, and task switching. Verbal interference does not appear to have a specific disruptive effect on visual change, visuospatial integration and wayfinding, reasoning with non-verbal materials, or theory of mind processing. For the reasoning with verbal materials and complex categorization categories, the evidence appears equivocal. Generally, the studies on reasoning with verbal materials that found a specific detrimental effect of verbal interference only found this effect when participants were highly skilled or trained (Gilhooly et al., 1999; Meiser et al., 2001) or when the premises were presented sequentially (Gilhooly et al., 2002). This might suggest that participants who had learned a strategy (probably through verbal instruction) were less able to use that under verbal interference conditions, and that inner speech was used to rehearse premises continuously to keep the memory of them from degrading. The studies on complex categorization that investigated novel category learning generally demonstrate involvement of working memory, but it remains somewhat unclear whether the verbal component of working memory plays a specific role (Maddox et al., 2004; Minda et al., 2008; Newell et al., 2010; Zeithamova & Maddox, 2007). The one study that tested complex categorization by abstracting over multiple categories did find a specific effect of verbal interference (Lupyan, 2009).

When does covert language use affect task performance?

Language appears to be recruited for solving problems by cuing yourself to remember the relevant task rule, naming shades of a color to distinguish it from other colors, or naming objects or features to be remembered. There is evidence of both implicit and spontaneous language effects and more explicit language strategies – our findings suggest people sometimes use very explicit verbal strategies to solve tasks, as seen for example in the context of reasoning with verbal materials. In general, it appears that covert language

Table 1 Primary task areas with evidence of covert language involvement. Note that some studies used multiple interference types and thus appear more than once in the "Interference task type" and "Specific effect of verbal interference" columns

Primary task area	Number of studies included in the review	Number of participants included in the review	Interference task type (N studies)	Specific effect of verbal interference (N/total studies)	Specific effect of verbal interference (N/total participants)
Categorization (complex)	5	982	Memory (4)	2/4	224/910
			Repetition (1)	1/1	72/72
Categorization (simple)	11	702	Memory (7)	5/7	362/401
			Judgment (1)	1/1	120/120
			Repetition (3)	1/3	135/181
Mental arithmetic	10	507	Memory (5)	3/5	185/353
			Repetition (4)	4/4	130/130
			Shadowing (1)	1/1	24/24
Memory	15	2110	Memory (2)	0/2	0/900
			Repetition (12)	10/12	918/1122
			Shadowing (1)	1/1	88/88
Motor control	2	50	Stroop task (2)	2/2	50/50
Reasoning (verbal materials)	8	900	Repetition (8)	4/8	696/900
Reasoning (non-verbal materials)	12	812	Repetition (9)	3/9	166/634
			Memory (5)	0/5	0/178
Task switching	16	1213	Repetition (16)	16/16	1213/1213
Theory of mind	4	243	Shadowing (3)	1/3	66/196
			Memory (1)	0/1	0/47
Visual change	6	248	Shadowing (3)	2/3	101/135
			Repetition (2)	1/2	12/27
			Same/different string (1)	0/1	0/86
Visuospatial integration and wayfinding	12	1126	Shadowing (7)	2/7	370/546
			Repetition (3)	0/3	0/364
			Word/non-word judgment (2)	0/2	0/216

aids cognition when the stimuli to be perceived, assessed, manipulated, or remembered lend themselves to a verbal code. We see this, for example, with the finding that naming objects makes them more likely to be remembered if names for their features exist, or with the finding that mental arithmetic problems demanding carry or borrow operations appear to be facilitated by language.

For categorization, the hypothesis is that covert language helps by providing a label to identify categories – this is an example of where the language effects appear to be implicit and involuntary. The fact that most of the studies reviewed indicated that verbal interference disrupts categorization fits well with the label-feedback hypothesis as proposed by Lupyan (e.g., 2012a, b). This hypothesis proposes that verbal labels – whether activated through overt or covert language use – feed-back on lower-level cognitive/perceptual processes with the effect of making them more categorical than they would be otherwise. In one study, Lupyan (2009) had participants judge which of three pictures (or words) was different from two others according either perceptual features (size, color), or more holistic thematic relationships. Under verbal interference, participants were worse at categorizing objects based on perceptual features but were still able to determine the odd one out based on thematic relationships – a pattern observed also in individuals with anomic aphasia (Cohen et al., 1980; Davidoff & Roberson, 2004; Lupyan & Mirman, 2013). Such results suggest that covert language is causally implicated in categorization tasks requiring isolation of specific dimensions (e.g., color). Recognizing that cherries and bricks, or snowmen and swans, have something in common is more difficult when language is interfered with or disrupted through a neurological insult. Additional support for this idea comes from studies using transcranial direct current stimulation (Lupyan et al., 2012; Perry & Lupyan, 2014), which have found that stimulating traditional language areas (left posterior superior temporal cortex, left inferior frontal cortex) disrupts the use of single-dimension categories.

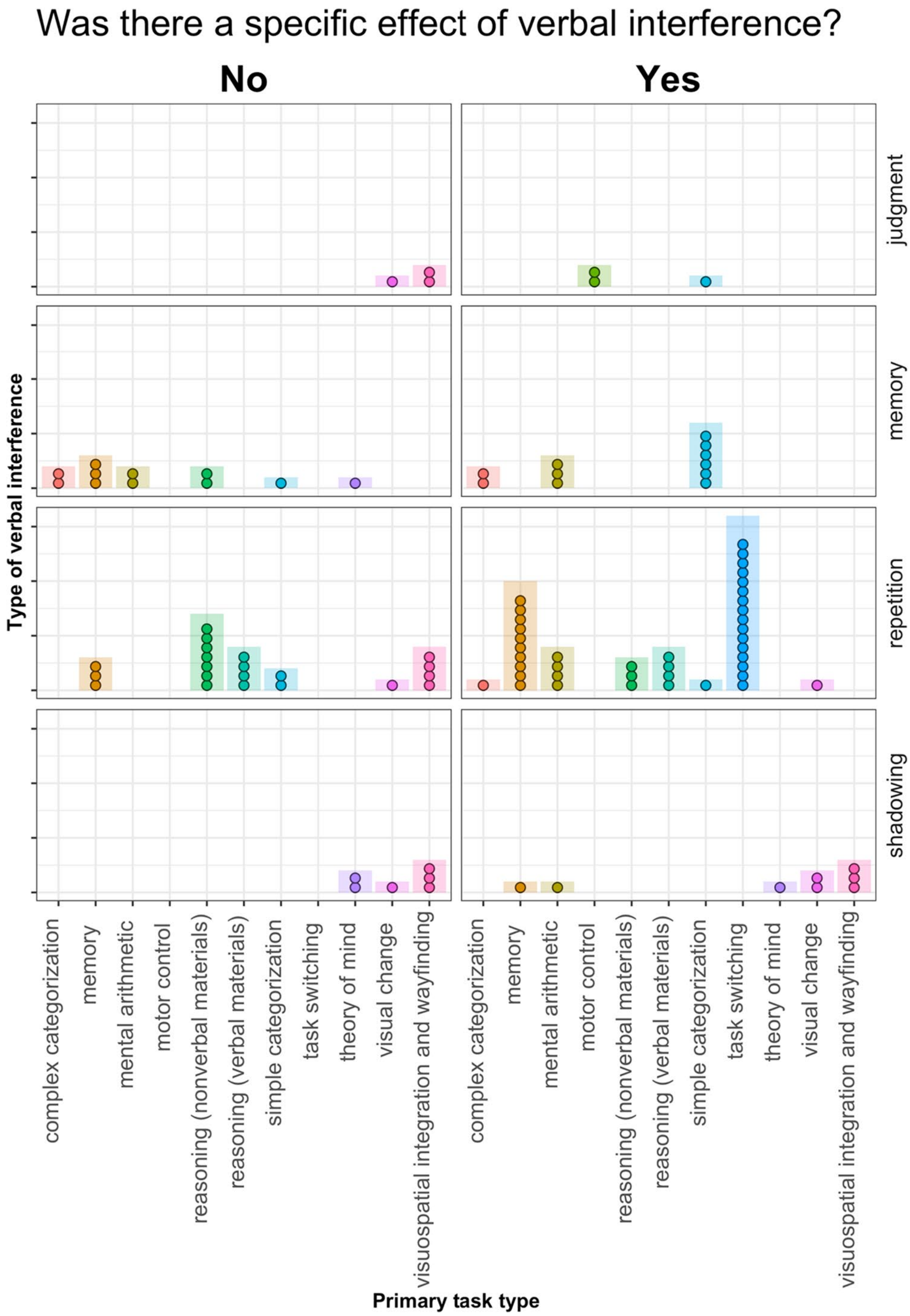


Fig. 2 Visualization of the overall results where each point represents a study included in the systematic review. The 11 primary task categories are indicated on the x axis and by color. Each row shows a different type of verbal interference. "Judgment" refers to judgment of verbal materials (for example rhyme), "memory" refers to the interference caused by a verbal memory task, "repetition" refers to repetition of simple syllables or words, and "shadowing" refers to the immediate repetition of continuously changing verbal material. Whether there was a specific effect of verbal interference (either compared with a non-verbal interference task or across different primary tasks) is indicated by the column-wise subplots in the plot grid. A version of this plot including sample sizes can be seen in the supplemental materials

Aside from isolating and abstracting over specific features for categorization, language also appears to be involved in discrimination and detection of already learned categories; Roberson and Davidoff (2000) investigated recognition memory for colors and facial expressions and found that verbal interference removed the advantage normally associated with categorical perception wherein cross-category judgments are more accurate than within-category judgments. Gilbert et al. (2006), (2008), Winawer et al. (2007), and He et al. (2019) all investigated color discrimination and found that there was a category advantage if the colors straddled color word boundaries and importantly that this effect disappeared with verbal interference. Roberson and Davidoff (2000) compared the effect of interference that used color words and non-color words, finding no difference between the two interference types. This suggests that the verbal interference effect they observed did not require cuing specifically task-relevant words. Interfering with language reduced categorical biases in color memory even when interference did not target color words. Converging evidence for effects of language on color memory comes from a study by Souza and Skóra (2017), who had participants remember colors while doing several tasks, among them, verbal interference and explicit color labeling (a form of up-regulation of language, see Perry & Lupyan, 2013). Unlike Roberson and Davidoff (2000), Souza and Skóra tested color memory by having participants select colors from a continuous distribution rather than through two-alternative forced choice. The authors found that explicit labeling decreased color memory in ways consistent with color labels inducing more categorical encoding in memory. Verbal interference during encoding did not affect color memory compared to control encoding conditions. A similar effect of explicit color-labeling increasing categoricality of color representations was found by Forder and Lupyan (2019), but this time on untimed color discrimination accuracy, rather than color memory.

Language does not just appear to affect cognition and perception by imposing labels and categories; however, there is also evidence that people use self-directed language to control their own behavior through rehearsal or self-cuing. In Emerson and Miyake's (2003) task-switching study, for example, verbal rehearsal plausibly helped maintain task set. This interpretation is supported by both the fact that the researchers found a specific effect of articulatory suppression and the fact that this effect depended on the existence of explicit cues to the relevant task. When there were explicit cues (plus and minus signs), articulatory suppression did not cause increased switch costs, indicating that the function of inner speech under no articulatory suppression is to provide these self-instruction cues. Asking participants to overtly verbalize the relevant cue to the task rule (presumably what they are doing covertly under normal circumstances), reduced response times, switching costs, and mixing costs (Goschke, 2000; Grange, 2013; Kirkham et al., 2012). In Nakabayashi and Burton (2008), participants were asked to remember faces – it is possible that covert language could be used as a mnemonic strategy in a similar way by allowing participants to verbalize specific features of the faces to be remembered (e.g., "potato nose," "high cheekbones," "no eyebrows", etc.) or an attempt to link faces to possible occupations or personalities. In fact, Nakabayashi and Burton (2008) found that when participants were asked to overtly describe the faces during learning, they were better at recognizing them than if they had just observed the faces silently, and Gimenes et al. (2016) found that training participants on a verbal strategy for remembering gestures improved their performance. In the four studies on reasoning with verbal materials that found specific effects of verbal interference (Farmer et al., 1986; Gilhooly et al., 1999, 2002; Meiser et al., 2001), the effects were only found for trained or highly skilled participants who had learned a specific strategy to solve the problems. As these strategies had been learned through verbal instruction, it is also likely that participants used inner speech to remind themselves of the relevant strategy for individual problems. It is also interesting that some studies found that disrupting verbal processing was associated with a loss of inhibitory control. For example, Dunbar and Sussman (1995) found that participants under verbal interference made more perseverative errors in the Wisconsin Card Sorting Task, Tullett and Inzlicht (2010) found that participants responded more impulsively on a Go/No-Go task, Wallace et al. (2017) found that participants made more excess moves on a Tower of London task while engaged in verbal interference, and both Biese et al. (2019) and Talarico et al. (2017) found that participants displayed poorer motor control while engaged in a simultaneous Stroop task.

Occasionally, effects of implicit labelling and overt strategies converge, as with nameability advantages of which there are many examples. Bek et al. (2009) investigated the Rey-Osterreith Complex Figure Test and the block design subtest of the Weschler Adult Intelligence Scale (in Experiment 1). They found that the block design task was unaffected by verbal shadowing, presumably because this task does not contain highly nameable features or require storage and rehearsal of visuospatial information. Contrastingly, copy and recall accuracy on the complex figure test were reduced if participants engaged in verbal shadowing during the copying stage and not if they were doing so during the recall stage. Verbal shadowing thus seemed to affect encoding rather than retrieval. The complex figure test notably had more nameable features than the block design test (e.g., "cross," "triangle") – participants are likely to have used these labels to support task performance and were prevented from doing so during shadowing. Further evidence for nameability advantages being sensitive to verbal interference comes from Walker and Cuthbert (1998), who investigated the unitization effect in color-shape associations. The unitization effect refers to the finding that memory for which visual properties occurred together is better if the properties are presented as belonging to the same object rather than separate objects (i.e., it is easier to remember a red triangle than a triangle *and* the color red). For our present purposes, the most interesting finding of this study was that the nameability advantage for particular shapes disappeared during articulatory suppression, suggesting that some kind of verbal recoding took place under normal circumstances. In a recent related study, Zettersten and Lupyan (2020) found that more nameable features improved rule-based category learning, although they did not find that this nameability effect was modulated by verbal interference.

In summary, it appears that language can aid cognition by providing labels for better memory and faster categorization, providing self-cues for self-control, task set reminders, and verbal strategies for problem solution, and by lending a medium for rehearsal or temporary storage of items in a verbal format (as with complex mental arithmetic). Importantly, it is not only overtly verbal strategies that appear to be interrupted by verbal interference but also more involuntary or spontaneous processes. This suggests that language can influence cognition beyond the surface level.

In what kinds of tasks does covert language *not* affect performance?

The present review found little support for the on-line role of covert language in various tasks relying on primarily visual processing (the categories we named visual change, visuospatial integration and wayfinding, and reasoning using non-verbal materials). To reiterate, the hypotheses for why language would be recruited for these tasks are that language is either necessary for integrating different kinds of features (e.g., color, shape, and locations) or that visuospatial stimuli are encoded both visually and linguistically, meaning that there is somehow weaker or more shallow processing if the verbal encoding is blocked. Judging by failures to replicate the results from Hermer-Vazquez et al. (1999), however, neither the former nor the latter putative roles are strongly supported. As for the other visually based tasks, the most plausible explanation is that solving the tasks efficiently requires participants to preserve a high degree of acuity with regard to the visual stimuli (maps, complex shapes, etc.), which rarely have nameability affordances. Thus, efficient and effective processing of the stimuli does not lend itself to a verbal code, and labelling specific aspects of the stimuli is not beneficial. Interrupting verbal processing is therefore not associated with a decrement in primary task performance.

The failure to find effects of verbal interference on performance in theory-of-mind-type tasks is interesting, especially as there is a large amount of evidence supporting the idea that language and theory of mind are intimately linked in development (Astington & Baird, 2005; Astington & Jenkins, 1999; Gagne & Coppola, 2017; Lohmann & Tomasello, 2003; Milligan et al., 2007; Pyers & Senghas, 2009; Slade & Ruffman, 2005). However, there is also evidence from adults with global aphasia suggesting that their theory-of-mind abilities are intact, which means that language and theory of mind are possibly only co-dependent during development (Siegal & Varley, 2006; Varley & Siegal, 2000). As previously discussed, there are two main theories on how language facilitates theory-of-mind development: either as a representational format providing the structure for representing mental states (i.e., sentential complements) or through directing children's attention to otherwise invisible mental state dynamics. Because the present review focused on adult participants, we cannot distinguish between these two theories. These apparently conflicting findings (that language and theory of mind appear to be linked in development but not in adult cognition) can potentially be resolved either by (a) language is recruited only for development and thus ceases to be necessary once theory of mind skills are acquired, or (b) the involvement of language and theory of mind has become so automatic and proceduralized in adults that verbal interference cannot affect it.

In some interesting cases, there was a specific effect of verbal interference, but this effect was not in the direction we expected. It is important to discuss these cases as it is often assumed that if language is recruited for cognition, this will always be in a facilitative way (Dove, 2020; Dove et al., 2020). In the memory studies, for example, verbal interference in several cases caused recognition memory to decrease while actually causing mental transformation performance to *increase* (Brandimonte et al., 1992a, b; Hitch et al., 1995; Pelizzon et al., 1999). The authors of these studies interpret this as meaning that we usually encode things to be remembered verbally but that encoding in this more abstract format actually makes visual encoding less detailed and thus less available for further manipulations. In a similar vein, verbal overshadowing research indicates that forcing verbal encoding of visual stimuli can cause memory performance to deteriorate (Alogna et al., 2014; Lane & Schooler, 2004; Schooler & Engstler-Schooler, 1990). In some additional cases, verbal interference also caused primary task processing to be faster (Evans & Brooks, 1981; Forgeot d'Arc & Ramus, 2011; Phillips, 1999), perhaps indicating that converting to a verbal code under normal circumstances takes time. It is also possible that verbal interference makes participants more likely to give their initial dominant response, which can cause more errors but faster responses.

It is important to note that a null result in a verbal interference experiment does not necessarily mean that language is in no way involved with that process. It is possible that language still affects the process but off-line, as, for example, discussed with regard to theory of mind where language looks to be involved during development, but not in on-line processing in adults. It is also possible that language is involved on-line but immune to verbal interference, for instance because its involvement has become so proceduralized and automatic that it can no longer be disrupted by superficial linguistic interference. This latter possibility is discussed in more detail by Wolff and Holmes (2011), who stated that 'the long-term use of a language may direct habitual attention to specific properties of the world, even in nonlinguistic contexts. At a more general level, language use may also induce a given mode of processing, which may persist even as people engage in other nonlinguistic tasks … these effects of 'thinking after language' should be less attenuated by verbal interference tasks, since they occur after language is no longer in use, rather than involving the recruitment of linguistic codes during processing.' (p. 259)

Choosing the interference task

It is a common problem that the different interference tasks are not matched in terms of general difficulty. One approach to this, taken by, for example, Lupyan (2009) and Hermer-Vazquez et al. (1999), is to check that the verbal and non-verbal interference tasks disrupt a third concurrent task to the same extent. This could for example be a visual search task. This approach is problematic, however, in that it glosses over the fact that the verbal and non-verbal components might also be differentially involved in this third concurrent task. It is difficult to choose a third concurrent task to validate the equivalence of the interference tasks because the literature is so divided on which tasks involve covert language and which do not. Another approach is to find a verbal and a non-verbal interference task that are in theory equivalent in every respect but their "verbality" (Perry & Lupyan, 2013), including performance. This approach faces challenges because tasks that are equivalent in everything but their verbality may yet place different demands on attention and executive function. Ideally, the tasks should at least be equated as separate single tasks in terms of their difficulty, and performance should neither be at ceiling nor at floor. This would make it possible to analyze potential trade-off effects with the primary task.

As we have seen, there are four types of verbal interference that have been used: syllable repetition, verbal memory, verbal shadowing, and judgment tasks. Only too rarely have the different interference tasks been directly compared, even though they might yield different predictions depending on which aspect of language (rehearsal, syntactic structure, verbal labels) you hypothesize is involved in the primary task you are investigating. Bek et al. (2009, 2013) directly compared syllable shadowing and prose shadowing, which should intuitively be different in terms of which components of language are involved. After all, syllable repetition uses less "language" than prose shadowing (semantics, syntax, morphology, etc.), which is precisely why syllable repetition is so widely used in working memory studies. In these experiments, there was no difference between shadowing syllables and shadowing prose. If anything, shadowing syllables resulted in a marginally more detrimental effect on visuospatial reorientation. A possible explanation may be that syllable shadowing lacks the predictability of prose shadowing and thus actually requires more cognitive resources.

Current forms of verbal interference (see above) are not well suited for distinguishing which components of language are most involved in performance on the primary task. Comparing interference involving task-relevant versus task-irrelevant words (Piccardi et al., 2020; Roberson & Davidoff, 2000) offers some, albeit limited, insights. A promising avenue for future research would be to compare manipulations designed to increase language involvement (e.g., as in Forder & Lupyan, 2019; Lupyan, 2008; Lupyan & Swingley, 2012) with conditions suppressing language involvement (e.g., as was done by Souza & Skóra, 2017). Once verbal interference has indicated that language in some form may be involved, up-regulating language involvement would be better suited to targeting specific hypotheses about components of language involved. We see this for example in findings indicating that the way language helps task switching is by helping to cue the relevant task rule (Goschke, 2000; Grange, 2013; Kirkham et al., 2012). Without additional task manipulations supplementing the dual-task interference, we would not have much indication as to *how* language helps task switching performance. Another example of up-regulating language shedding light on the specific ways language may be involved comes from the sport psychology literature where self-talk interventions (up-regulating language) are much more common than dual-task interference studies (Hatzigeorgiadis et al., 2011; Tod et al., 2011). Here, participants are often trained to use different types of self-directed verbalizations (instructional vs. motivational, positive vs. negative, etc.), which result in different effects on performance depending on the participant's skill level (Zourbanos et al., 2013), the motor demands of the sport (Theodorakis et al., 2000), and whether the self-talk takes place in a competition or practice context (Hatzigeorgiadis et al., 2014). In addition to focusing on the content of internal verbalizations, it is also important to understand the *stage* at which interfering with language affects performance, for example, during memory encoding, retrieval, or both (Frank et al., 2012; Nakabayashi & Burton, 2008). This may help tease apart effects of verbal encoding (nameability effects in memory, verbal overshadowing) and "mental workspace" functions (using the phonological loop to keep track of carry or borrow operations, keeping track of the relevant task rule). Future studies would benefit from clarifying their predictions about language involvement in this way.

Summary of suggestions for future studies

Future studies should follow these recommendations:

1. Include control conditions of both the primary and the secondary tasks.
2. Make theoretically informed and hypothesis-driven choices about the type of interference task and/or directly compare effects of different types.
3. Ensure that the different interference tasks are matched in terms of difficulty/attentional demands by measuring performance.
4. Consider potential trade-offs between effort/resources put into the primary tasks and the secondary tasks.
5. Delineate the precise mechanisms by which language is expected to help cognition.

Conclusion

It appears that language – including inner speech – is a powerful tool for directing attention, improving memory, and controlling actions. These three processes, however, are intimately connected. For example, paying attention to specific aspects or properties of something makes it more likely that you will remember it later, and remembering how you acted in a past situations can (and should) influence what you attend to and how you act in the current situation. We reviewed 101 studies investigating the on-line role of language in some cognitive function using a dual-task interference methodology. Overall, we found that it is likely the case that covert language is recruited for behavioral self-cuing (inhibitory control, task set reminders, verbal strategy), rehearsal for memory when items to be remembered have readily available labels, and as a workspace for complex mental arithmetic. We found less evidence for a role of on-line language use in cross-modal integration, reasoning that relies on a high degree of visual detail (such as map tasks, visual recursion tasks, and some matrix problems), and theory of mind. It is important to note that we only examined *one* way of investigating the role of language in cognition and that other patterns of effects may appear with the use of different approaches. Interestingly, we found that recruiting language for non-verbal tasks is not always purely advantageous, but may present costs in term of processing speed, loss of visual detail, and verbal overshadowing. Future studies should include relevant control conditions for both primary and secondary tasks, make informed and justified decisions about the interference tasks, ensure that the interference tasks are appropriately matched, and delineate the precise mechanisms by which covert language is expected to help cognition in the on-line processing of a given primary task.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-022-02144-7.

Declarations

This work was supported by a PhD grant awarded to the first author. The authors all declare no conflicts of interest. The full table of the papers included in this review can be found in Appendix A, which is available as Online Supplementary Material.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*(5), 931–965. https://doi.org/10.1037/bul0000021

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., Brown, C., Buswell, K., Carlson, C., Carlson, M., Chu, S., Cislak, A., Colarusso, M., Colloff, M. F., Dellapaolera, K. S., Delvenne, J.-F., … Zwaan, R. A. (2014). Registered Replication Report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578. https://doi.org/10.1177/1745691614545653

Annett, J. M., & Leslie, J. C. (1996). Effects of visual and verbal interference tasks on olfactory memory: The role of task complexity. *British Journal of Psychology, 87*(3), 447–460. https://doi.org/10.1111/j.2044-8295.1996.tb02601.x

Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition, 1*(1), 3–34.

Astington, J. W., & Baird, J. A. (2005). *Why language matters for theory of mind*. Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311.

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. https://doi.org/10.1080/14640746608400055

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology, 60*(4), 497–504. https://doi.org/10.1080/17470210601147572

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), 641–657. https://doi.org/10.1037/0096-3445.130.4.641

Baldo, J., Dronkers, N., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language, 92*(3), 240–250. https://doi.org/10.1016/j.bandl.2004.06.103

Bek, J., Blades, M., Siegal, M., & Varley, R. (2009). Linguistic processes in visuospatial representation: Clarifying verbal interference effects. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*.

Bek, J., Blades, M., Siegal, M., & Varley, R. (2013). Dual-task interference in spatial reorientation: Linguistic and nonlinguistic factors. *Spatial Cognition & Computation, 13*(1), 26–49. https://doi.org/10.1080/13875868.2011.590622

Bermúdez, J. L. (2003). *Thinking without words*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159691.001.0001

Biese, K. M., Pietrosimone, L. E., Andrejchak, M., Lynall, R. C., Wikstrom, E. A., & Padua, D. A. (2019). Preliminary investigation on the effect of cognition on jump-landing performance using a clinically relevant setup. *Measurement in Physical Education and Exercise Science, 23*(1), 78–88. https://doi.org/10.1080/1091367X.2018.1518875

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 157–165. https://doi.org/10.1037/0278-7393.18.1.157

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual stimuli impairs mentalimagetransformations. *Memory & Cognition, 20*(4), 449–455. https://doi.org/10.3758/BF03210929

Brown, R. G., & Marsden, C. D. (1991). Dual task performance and processing resources in normal subjects and patients with Parkinson's disease. *Brain, 114A*(1), 215–231. https://doi.org/10.1093/oxfordjournals.brain.a101858

Bryck, R. L., & Mayr, U. (2005). On the role of verbalization during task set selection: Switching or serial order control? *Memory & Cognition, 33*(4), 611–623. https://doi.org/10.3758/BF03195328

Caffò, A. O., Picucci, L., Di Masi, M. N., & Bosco, A. (2011). Working memory components and virtual reorientation: A dual-task study. In E. S. Levin (Ed.), *Working memory: Capacity, developments and improvement techniques* (pp. 249–266). Nova Science Publisher.

Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *Quarterly Journal of Experimental Psychology, 66*(10), 1927–1950. https://doi.org/10.1080/17470218.2013.768275

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*(6), 657–674. https://doi.org/10.1017/S0140525X02000122

Cheetham, J. M., Rahm, B., Kaller, C. P., & Unterrainer, J. M. (2012). Visuospatial over verbal demands in predicting Tower of London planning tasks: Visuospatial demands during TOL planning tasks. *British Journal of Psychology, 103*(1), 98–116. https://doi.org/10.1111/j.2044-8295.2011.02049.x

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought* (1st ed., pp. 162–183). Cambridge University Press. https://doi.org/10.1017/CBO9780511597909.011

Clearman, J., Klinger, V., & Szűcs, D. (2017). Visuospatial and verbal memory in mental arithmetic. *Quarterly Journal of Experimental Psychology, 70*(9), 1837–1855. https://doi.org/10.1080/17470218.2016.1209534

Coetzee, J., Monti, M., Iacoboni, M., Wu, A., & Johnson, M. (2019). Separability of logic and language: A TMS study. *Brain Stimulation, 12*(2), 543. https://doi.org/10.1016/j.brs.2018.12.792

Cohen, R., Kelter, S., & Woll, G. (1980). Analytical competence and language impairment in aphasia. *Brain and Language, 10*(2), 331–347. https://doi.org/10.1016/0093-934X(80)90060-7

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55*, 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55*, 429–432.

Croijmans, I., Arshamian, A., Speed, L. J., & Majid, A. (2021). Wine experts' recognition of wine odors is not verbally mediated. *Journal of Experimental Psychology: General, 150*(3), 545–559. https://doi.org/10.1037/xge0000949

Davidoff, J., & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes, 19*(1), 137–174. https://doi.org/10.1080/01690960344000125

De Renzi, E., & Spinnler, H. (1967). Impaired performance on color tasks in patients with hemispheric damage. *Cortex, 3*(2), 194–217. https://doi.org/10.1016/S0010-9452(67)80012-1

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858–1878.

de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Psychology Press.

de Villiers, J., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1), 1037–1060.

Dent, K. (2009). Coding categorical and coordinate spatial relations in visual–spatial short-term memory. *Quarterly Journal of Experimental Psychology, 62*(12), 2372–2387. https://doi.org/10.1080/17470210902853548

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science, 19*(2), 189–195. https://doi.org/10.1111/j.1467-9280.2008.02066.x

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353–386. https://doi.org/10.1080/09541440244000328

Dove, G. (2020). More than a scaffold: Language is a neuroenhancement. *Cognitive Neuropsychology, 37*(5–6), 288–311.

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2020). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research*, 1–17.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 6*(5), 178–190.

Dunbar, K., & Sussman, D. (1995). Toward a cognitive account of frontal lobe function: Simulating frontal lobe deficits in normal subjects. *Annals of the New York Academy of Sciences, 769*(1 Structure and), 289–304. https://doi.org/10.1111/j.1749-6632.1995.tb38146.x

Dungan, J., & Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cognitive Science, 36*(6), 1148–1156. https://doi.org/10.1111/j.1551-6709.2012.01248.x

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. https://doi.org/10.1016/S0749-596X(02)00511-9

Evans, J. S. B. T., & Brooks, P. G. (1981). Competing with reasoning: A test of the working memory hypothesis. *Current Psychology, 1*(2), 139–147. https://doi.org/10.1007/BF02684486

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology Section A, 38*(4), 675–688. https://doi.org/10.1080/14640748608401620

Feinmann, D. (2020). Language and thought in the motion domain: Methodological considerations and new empirical evidence. *Journal of Psycholinguistic Research, 49*(1), 1–29. https://doi.org/10.1007/s10936-019-09668-5

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

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. https://doi.org/10.1037/xge0000560

Forgeot d'Arc, B., & Ramus, F. (2011). Belief attribution despite verbal interference. *Quarterly Journal of Experimental Psychology, 64*(5), 975–990. https://doi.org/10.1080/17470218.2010.524413

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology, 64*(1–2), 74–92. https://doi.org/10.1016/j.cogpsych.2011.10.004

Gagne, D. L., & Coppola, M. (2017). Visible social interactions do not support the development of false belief understanding in the absence of linguistic input: Evidence from deaf adult homesigners. *Frontiers in Psychology, 8,* 837. https://doi.org/10.3389/fpsyg.2017.00837

Gaillard, V., Destrebecqz, A., & Cleeremans, A. (2012). The influence of articulatory suppression on the control of implicit sequence knowledge. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00208

Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology, 16*(1), 35–50. https://doi.org/10.1002/acp.746

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. https://doi.org/10.1073/pnas.0509868103

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. https://doi.org/10.1016/j.bandl.2007.06.001

Gilhooly, K. J. (2005). Working memory and strategies in reasoning. In E. Newton & M. Roberts (Eds.), *Methods of thought: Individual differences in reasoning strategies* (1st ed., pp. 57–80).

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*(1), 115–124. https://doi.org/10.3758/BF03211170

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11*(4), 473–498. https://doi.org/10.1080/095414499382264

Gilhooly, K. J., Wynn, V., Phillips, L. H., Logie, R. H., & Sala, S. D. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking & Reasoning, 8*(3), 165–178. https://doi.org/10.1080/13546780244000006

Gimenes, G., Pennequin, V., & Mercer, T. (2016). What is the best strategy for retaining gestures in working memory? *Memory, 24*(6), 757–765. https://doi.org/10.1080/09658211.2015.1049544

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge University Press.

Gomila, A., Travieso, D., & Lobo, L. (2012). Wherein is human cognition systematic? *Minds and Machines, 22*(2), 101–115. https://doi.org/10.1007/s11023-012-9277-z

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Attention & performance XVIII: Control of cognitive processes*. MIT Press.

Grange, J. A. (2013). Manipulating task preparation via verbalisations does not affect inhibition in set switching. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2350387

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., & Theodorakis, Y. (2011). Self-talk and sports performance: A meta-analysis. *Perspectives on Psychological Science, 6*(4), 348–356. https://doi.org/10.1177/1745691611413136

Hatzigeorgiadis, A., Galanis, E., Zourbanos, N., & Theodorakis, Y. (2014). Self-talk and Competitive Sport Performance. *Journal of Applied Sport Psychology, 26*(1), 82–95. https://doi.org/10.1080/10413200.2013.790095

He, H., Li, J., Xiao, Q., Jiang, S., Yang, Y., & Zhi, S. (2019). Language and color perception: Evidence from Mongolian and Chinese speakers. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00551

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition, 28*(3), 376–385. https://doi.org/10.3758/BF03198553

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A, 56*(8), 1307–1334. https://doi.org/10.1080/02724980244000747

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: dual-task studies of space and language. *Cognitive Psychology, 39*(1), 3–36. https://doi.org/10.1006/cogp.1998.0713

Hintzman, D. L. (1967). Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6*(3), 312–316.

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination. *Memory & Cognition, 23*(2), 147–154. https://doi.org/10.3758/BF03197217

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 388–403. https://doi.org/10.1037/0096-1523.29.2.388

Hund, A. M. (2016). Visuospatial working memory facilitates indoor wayfinding and direction giving. *Journal of Environmental Psychology, 45,* 233–238. https://doi.org/10.1016/j.jenvp.2016.01.008

Hupbach, A., Hardt, O., Nadel, L., & Bohbot, V. D. (2007). Spatial reorientation: Effects of verbal and spatial shadowing. *Spatial Cognition & Computation, 7*(2), 213–226. https://doi.org/10.1080/13875860701418206

Imbo, I., & LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese- and Canadian-educated adults. *Memory & Cognition, 38*(2), 176–185. https://doi.org/10.3758/MC.38.2.176

Imbo, I., De Rammelaere, S., & Vandierendonck, A. (2005). New insights in the role of working memory in carry and borrow operations. *Psychologica Belgica, 45*(2), 101–121. https://doi.org/10.5334/pb-45-2-101

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 656–674.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology, 60*(4), 505–511. https://doi.org/10.1080/17470210601147598

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849–874. https://doi.org/10.1037/a0019842

Kirkham, A. J., Breeze, J. M., & Marí-Beffa, P. (2012). The impact of verbal instructions on goal-directed behaviour. *Acta Psychologica, 139*(1), 212–219. https://doi.org/10.1016/j.actpsy.2011.09.016

Klauer, K. C. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning, 3*(1), 9–47. https://doi.org/10.1080/135467897394419

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557–583. https://doi.org/10.1037/bul0000144

Kranjec, A., Lupyan, G., & Chatterjee, A. (2014). Categorical Biases in Perceiving Spatial Relations. *PLoS ONE, 9*(5), e98604. https://doi.org/10.1371/journal.pone.0098604

Labate, E., Pazzaglia, F., & Hegarty, M. (2014). What working memory subcomponents are needed in the acquisition of survey knowledge? Evidence from direction estimation and shortcut tasks. *Journal of Environmental Psychology, 37,* 73–79. https://doi.org/10.1016/j.jenvp.2013.11.007

Lane, S. M., & Schooler, J. W. (2004). Skimming the surface: Verbal overshadowing of analogical retrieval. *Psychological Science, 15*(11), 715–719. https://doi.org/10.1111/j.0956-7976.2004.00747.x

Law, A. S., Logie, R. H., & Pearson, D. G. (2006). The impact of secondary tasks on multitasking in a virtual environment. *Acta Psychologica, 122*(1), 27–44. https://doi.org/10.1016/j.actpsy.2005.09.002

Law, A. S., Trawley, S. L., Brown, L. A., Stephens, A. N., & Logie, R. H. (2013). The impact of working memory load on task execution and online plan adjustment during multitasking in a virtual environment. *Quarterly Journal of Experimental Psychology, 66*(6), 1241–1258. https://doi.org/10.1080/17470218.2012.748813

Lee, K.-M., & Kang, S.-Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition, 83*(3), B63–B68. https://doi.org/10.1016/S0010-0277(02)00010-0

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology, 58*(4), 487–524. https://doi.org/10.1016/j.cogpsych.2008.12.001

Liefooghe, B., Vandierendonck, A., Muyllaert, I., Verbruggen, F., & Vanneste, W. (2005). The phonological loop in task alternation and task repetition. *Memory, 13*(5), 550–560. https://doi.org/10.1080/09658210444000250

Liu, Q., Chen, A.-T., Wang, Q., Zhou, L., & Sun, H.-J. (2008). An evidence for the effect of categorical perception on color perception: An evidence for the effect of categorical perception on color perception. *Acta Psychologica Sinica, 40*(1), 8–13. https://doi.org/10.3724/SP.J.1041.2008.00008

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*(4), 395–410. https://doi.org/10.3758/BF03200866

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130–1144. https://doi.org/10.1111/1467-8624.00597

Luo, X., Sexton, N. J., & Love, B. C. (2021). A deep learning account of how language affects thought. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2021.2001023

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*(2), 566–577.

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review, 16*(4), 711–718. https://doi.org/10.3758/PBR.16.4.711

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Academic Press. http://www.sciencedirect.com/science/article/pii/B9780123942937000078

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex, 49*(5), 1187–1194. https://doi.org/10.1016/j.cortex.2012.06.006

Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology, 65*(6), 1068–1085. https://doi.org/10.1080/17470218.2011.647039

Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition, 124*(1), 36–49. https://doi.org/10.1016/j.cognition.2012.04.002

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*(3), 309–332. https://doi.org/10.1016/j.beproc.2004.03.011

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition, 32*(4), 582–591. https://doi.org/10.3758/BF03195849

Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 265–284). John Benjamins Publishing Company.

Martins, M. de J. D., Muršič, Z., Oh, J., & Fitch, W. T. (2015). Representing visual recursion does not require verbal or motor resources. *Cognitive Psychology, 77,* 20–41. https://doi.org/10.1016/j.cogpsych.2015.01.004

Meilinger, T., Knauff, M., & Bülthoff, H. H. (2008). Working memory in wayfinding-a dual task experiment in a virtual City. *Cognitive Science, 32*(4), 755–770. https://doi.org/10.1080/03640210802067004

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content q. *Acta Psychologica, 25*.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1518–1533. https://doi.org/10.1037/a0013355

Mitsuhashi, S., Hirata, S., & Okuzumi, H. (2018). Role of inner speech on the Luria hand test. *Cogent Psychology, 5*(1), 1449485. https://doi.org/10.1080/23311908.2018.1449485

Miyake, A., Witzki, A. H., & Emerson, M. J. (2001). Field dependence–independence from a working memory perspective: A dual-task investigation of the Hidden Figures Test. *Memory, 9*(4–6), 445–457. https://doi.org/10.1080/09658210143000029

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. https://doi.org/10.1016/j.actpsy.2003.12.004

Müller, M. (1978). The science of thought. In *The science of thought /* (Repr.). AMS Press.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 21*(3), 263–276.

Nakabayashi, K., & Burton, A. M. (2008). The role of verbal processing at different stages of recognition memory for faces. *European Journal of Cognitive Psychology, 20*(3), 478–496. https://doi.org/10.1080/09541440801946174

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition, 38*(5), 563–581. https://doi.org/10.3758/MC.38.5.563

Newton, A. M., & de Villiers, J. G. (2007). Thinking while talking: Adults fail nonverbal false-belief reasoning. *Psychological Science, 18*(7), 574–579. https://doi.org/10.1111/j.1467-9280.2007.01942.x

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*(3), 255–287. https://doi.org/10.1037/h0084295

Pelizzon, L., Brandimonte, M. A., & Favretto, A. (1999). Imagery and recognition: Dissociable measures of memory? *European Journal of Cognitive Psychology, 11*(3), 429–443. https://doi.org/10.1080/713752323

Perkins, J., & McLaughlin Cook, N. (1990). Recognition and recall of odours: The effects of suppressing visual and verbal encoding processes. *British Journal of Psychology, 81*(2), 221–226. https://doi.org/10.1111/j.2044-8295.1990.tb02357.x

Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience, 7*. https://doi.org/10.3389/fnbeh.2013.00122

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. https://doi.org/10.1016/j.bandl.2014.05.005

Peterson, L. R. (1969). Concurrent verbal activity. *Psychological Review, 76*(4), 376–386.

Phillips, L. H. (1999). The role of memory in the tower of London task. *Memory, 7*(2), 209–231. https://doi.org/10.1080/741944066

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the tower of London task. *Memory, 7*(2), 209–231. https://doi.org/10.1080/741944066

Piccardi, L., Bocchi, A., Palmiero, M., Boccia, M., D'Amico, S., & Nori, R. (2020). Chatting while walking does not interfere with topographical working memory. *Brain Sciences, 10*(11), 811. https://doi.org/10.3390/brainsci10110811

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition, 31*(4), 538–551. https://doi.org/10.3758/BF03196095

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychological Science, 20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Rao, K. V., & Baddeley, A. (2013). Raven's matrices and working memory: A dual-task approach. *Quarterly Journal of Experimental Psychology, 66*(10), 1881–1887. https://doi.org/10.1080/17470218.2013.828314

Ratliff, K. R., & Newcombe, N. S. (2005). *Human spatial reorientation using dual task paradigms* (pp. 6).

Ratliff, K. R., & Newcombe, N. S. (2008). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology, 56*(2), 142–163. https://doi.org/10.1016/j.cogpsych.2007.06.002

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*(6), 977–986. https://doi.org/10.3758/BF03209345

Robert, N. D., & LeFevre, J.-A. (2013). Ending up with less: The role of working memory in solving simple subtraction problems with positive and negative answers. *Research in Mathematics Education, 15*(2), 165–176. https://doi.org/10.1080/14794802.2013.797748

Romano, S., Salles, A., Amalric, M., Dehaene, S., Sigman, M., & Figueira, S. (2018). Bayesian validation of grammar productions for the language of thought. *PLOS ONE, 13*(7), e0200420. https://doi.org/10.1371/journal.pone.0200420

Saeki, E. (2007). Phonological loop and goal maintenance: effect of articulatory suppression in number-size consistency task. *Psychologia, 50*(2), 122–131. https://doi.org/10.2117/psysoc.2007.122

Saeki, E., & Saito, S. (2004a). The role of the phonological loop in task switching performance: The effect of articulatory suppression in the alternating runs paradigm. *Psychologia, 47*(1), 35–43. https://doi.org/10.2117/psysoc.2004.35

Saeki, E., & Saito, S. (2004b). Effect of articulatory suppression on task-switching performance: Implications for models of working memory. *Memory, 12*(3), 257–271. https://doi.org/10.1080/09658210244000649

Saeki, E., & Saito, S. (2009). Verbal representation in task order control: An examination with transition and task cues in random task switching. *Memory & Cognition, 37*(7), 1040–1050. https://doi.org/10.3758/MC.37.7.1040

Saeki, E., Saito, S., & Kawaguchi, J. (2006). Effects of response–stimulus interval manipulation and articulatory suppression on task switching. *Memory, 14*(8), 965–976. https://doi.org/10.1080/09658210601008973

Saeki, E., Baddeley, A., Hitch, G. J., & Saito, S. (2013). Breaking a habit: A further role of the phonological loop in action control. *Memory & Cognition, 41*(7), 1065–1078. https://doi.org/10.3758/s13421-013-0320-y

Samuel, S., Durdevic, K., Legg, E. W., Lurz, R., & Clayton, N. S. (2019). Is language required to represent others' mental states? Evidence from beliefs and other representations. *Cognitive Science, 43*(1), e12710. https://doi.org/10.1111/cogs.12710

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*(1), 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Seitz, K., & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology, 12*(4), 552–570. https://doi.org/10.1080/095414400750050231

Seitz, K., & Schumann-Hengsteler, R. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychologische Beiträge, 44*(2), 275.

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods, 49*(3), 853–862. https://doi.org/10.3758/s13428-016-0741-1

Siegal, M., & Varley, R. (2006). Aphasia, language, and theory of mind. *Social Neuroscience, 1*(3–4), 167–174.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7*(5), 301–305. https://doi.org/10.1111/j.1467-9280.1996.tb00378.x

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition, 25*(3), 321–332. https://doi.org/10.3758/BF03211288

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*(1), 117–141. https://doi.org/10.1348/026151004X21332

Sokolov, A. N. (1968). *Inner speech and thought* (G. T. Onischenko, Trans.). Plenum Press.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition, 166,* 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Talarico, M. K., Lynall, R. C., Mauntel, T. C., Weinhold, P. S., Padua, D. A., & Mihalik, J. P. (2017). Static and dynamic single leg postural control performance during dual-task paradigms. *Journal of Sports Sciences, 35*(11), 1118–1124. https://doi.org/10.1080/02640414.2016.1211307

Theodorakis, Y., Weinberg, R., Natsis, P., Douma, I., & Kazakas, P. (2000). The effects of motivational versus instructional self-talk on improving motor performance. *The Sport Psychologist, 14*(3), 253–271.

Tod, D., Hardy, J., & Oliver, E. (2011). Effects of self-talk: A systematic review. *Journal of Sport and Exercise Psychology, 33*(5), 666–687. https://doi.org/10.1123/jsep.33.5.666

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology Section A, 46*(4), 679–699. https://doi.org/10.1080/14640749308401033

Trbovich, P. L., & LeFevre, J.-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*(5), 738–745. https://doi.org/10.3758/BF03196112

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica, 135*(2), 252–256. https://doi.org/10.1016/j.actpsy.2010.07.008

van der Ham, I. J. M., & Borst, G. (2011). The nature of categorical and coordinate spatial relation processing: An interference study. *Journal of Cognitive Psychology, 23*(8), 922–930. https://doi.org/10.1080/20445911.2011.586780

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*(1), 57–79. https://doi.org/10.1348/000712604322779460

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind'in an agrammatic aphasic patient. *Current Biology, 10*(12), 723–726.

Vignolo, L. A. (1999). Disorders of conceptual thinking in aphasia. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology* (p. 273e288). Psychology Press.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114. https://doi.org/10.1037/0096-1523.27.1.92

Vygotsky, L. (1962). Thought and language. In *Thought and language /*. Massachusetts Institute of Technology. John Wiley.

Walker, P., & Cuthbert, L. (1998). Remembering visual feature conjunctions: Visual memory for shape-colour associations is object-based. *Visual Cognition, 5*(4), 409–455. https://doi.org/10.1080/713756794

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. https://doi.org/10.1044/2017_JSLHR-S-16-0376

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28*(7), 1205–1212. https://doi.org/10.3758/BF03211821

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*(2), 158–177. https://doi.org/10.1037/h0074428

Weywadt, C. R. B., & Butler, K. M. (2013). The role of verbal short-term memory in task selection: How articulatory suppression influences task choice in voluntary task switching. *Psychonomic Bulletin & Review, 20*(2), 334–340. https://doi.org/10.3758/s13423-012-0349-0

Wickelgren, W. A. (1965a). Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 4*(1), 53–61. https://doi.org/10.1016/S0022-5371(65)80067-6

Wickelgren, W. A. (1965b). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70*(1), 102–108. https://doi.org/10.1037/h0022015

Wickham, L. H. V., & Swift, H. (2006). Articulatory suppression attenuates the verbal overshadowing effect: A role for verbal encoding in face identification. *Applied Cognitive Psychology, 20*(2), 157–169. https://doi.org/10.1002/acp.1176

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. https://doi.org/10.1073/pnas.0701644104

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision, 11*(12), 16. https://doi.org/10.1167/11.12.16

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science, 2*(3), 253–265. https://doi.org/10.1002/wcs.104

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*(6), 1380–1398. https://doi.org/10.3758/BF03193609

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition, 196,* 104135. https://doi.org/10.1016/j.cognition.2019.104135

Zhong, W., Li, Y., Li, P., Xu, G., & Mo, L. (2015). Short-term trained lexical categories produce preattentive categorical perception of color: Evidence from ERPs: Categories produce categorical perception. *Psychophysiology, 52*(1), 98–106. https://doi.org/10.1111/psyp.12294

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences, 107*(22), 9974–9978. https://doi.org/10.1073/pnas.1005669107

Zourbanos, N., Hatzigeorgiadis, A., Bardas, D., & Theodorakis, Y. (2013). The effects of self-talk on dominant and nondominant arm performance on a handball task in primary physical education students. *The Sport Psychologist, 27*(2), 171–176. https://doi.org/10.1123/tsp.27.2.171

Article II

Mind over body: Interfering with the inner voice is detrimental to endurance performance

Johanne Nedergaard[1], Mark Schram Christensen[2,3], & Mikkel Wallentin[1]

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

[2]Department of Psychology, Cognition, Intention and Action group, University of Copenhagen

[3]Department of Neuroscience, Christensen Lab, University of Copenhagen

Author note

All authors declare that they have no competing interests. All experiment data and code can be accessed at https://osf.io/uk2y4/?view_only=6fc8f12830df497e9c403cfb01ebc66c. The two preregistrations can be accessed at https://osf.io/2ah7s and https://osf.io/byfp3.

Correspondence concerning this article should be addressed to Johanne Nedergaard, Jens Chr Skous Vej 2, 8200 Aarhus N, Denmark. E-mail: nedergaard@cc.au.dk

Abstract

In two preregistered experiments, we investigated whether covert language is involved in motivation, specifically if people are less able to push themselves physically when distracted from using inner speech. In both experiments, participants performed 12 cycling trials (Experiment 1: N = 49; Experiment 2: N = 50), each lasting one minute where participants were required to cycle as fast as possible while simultaneously engaging in either a visuospatial task, a verbal task or no interference. Experiment 1: Participants performed worse in the verbal interference condition compared with the control condition ($d = 0.29$) and verbal interference performance was numerically worse than visuospatial interference ($d = 0.22$). Experiment 2: A more demanding interference task yielded significant slower cycling with verbal interference compared to both control ($d = 1$) and spatial interference ($d = 0.43$). These results indicate that inner speech plays a causal role in control of sustained physical efforts.

Keywords: dual-task interference; cognitive control; self-regulation; endurance; inner speech

Word count: 7942

1. INTRODUCTION

Language and motor control are usually conceived of as separate cognitive systems with little influence on each other. However, if we consider (prolonged, sustained) motor control as requiring executive functions, then a connection seems plausible, as executive functions have a long, linked history with language (Alderson-Day & Fernyhough, 2015b; Cragg & Nation, 2010). Covert language plays a role in cognitive control (Baddeley et al., 2001; Bryck & Mayr, 2005; Emerson & Miyake, 2003; Tullett & Inzlicht, 2010; G. L. Wallace et al., 2017), and cognitive control is required for optimal physical performance (Brick et al., 2016; Hyland-Monks et al., 2018; Kirschenbaum, 1987; McCormick et al., 2019). In the present study, we combine findings from sport psychology and methods from cognitive psychology: In sport psychology, self-talk has been found to improve performance while dual-task interference paradigms from cognitive psychology have been used to investigate the role of verbal rehearsal in various cognitive processes. We conducted two experiments testing non-expert participants' cycling performance on an exercise bike under two different interference conditions (verbal and visuospatial) and a no-interference control condition. This extends current findings by testing a causal link between inner speech and endurance performance and by applying the dual-task method to a novel area of top-down control.

1.1. Verbal rehearsal and cognitive control

The core executive functions include inhibition, interference control, working memory, and cognitive flexibility (Diamond, 2013). Covert language is involved in several executive processes as people use it to control their own behaviour and remind themselves what their task is (Baddeley et al., 2001; Baldo et al., 2005; Dunbar & Sussman, 1995; Emerson & Miyake, 2003; Henson et al., 2003; Tullett & Inzlicht, 2010). These findings are based on the dual-task interference method where participants are asked to perform a primary task (e.g., adding and subtracting numbers) while also performing concurrent interference task (e.g., repeating the word 'the'). Using this method, inner speech has for example been shown to be involved in impulsivity and inhibitive control (Baldo et al., 2005; Dunbar & Sussman, 1995). Tullett and Inzlicht (2010) tested a go/no go task under verbal (repeating the word "computer" at 2 Hz) and spatial (drawing circles) interference conditions and found that verbal interference increased impulsive responding (faster responses, more commission errors, fewer omission errors). There is also evidence to suggest that people use inner speech to cue themselves on what the relevant task is if they have to switch between multiple task rules (Baddeley et al., 2001; Emerson & Miyake, 2003). While the dual-task method has been a very popular tool for testing the role of covert language in various tasks, it has not yet been used specifically to investigate the role of inner speech in motor control (Nedergaard et al., 2022).

Translating these findings to the area of sustained, physical effort, we expect cycling performance to be related to inhibitory control (or response inhibition) via the ability to resist temptations and to resist acting impulsively. In the case of endurance cycling, the impulse to be inhibited is the impulse to stop when the physical exertion becomes uncomfortable. We might also expect inner speech to play a role in sustained, physical effort through behavioural self-cuing whereby participants focus their own attention on the task instead of allowing it to drift away.

1.2. Self-talk in sport psychology

Self-talk interventions generally have positive effects on performance across a range of sports (Tod et al., 2011; Hatzigeorgiadis et al., 2011). It is a robust finding that (especially endurance) athletes use self-talk to a very large extent (Van Raalte et al., 2015) and that they believe it helps them perform better (Nedergaard et al., 2021). It is, however, still an open question whether inner speech in fact helps control physical performance beyond what athletes believe. Only a few studies to date have directly investigated self-talk in endurance sport through interventions where participants are typically trained to use specific self-talk phrases (Barwood et al., 2015; Blanchfield et al., 2014; Hamilton et al., 2007; Hatzigeorgiadis et al., 2018; McCormick et al., 2018; Schüler &

Langens, 2007; P. J. Wallace et al., 2017). These intervention studies, therefore, do not typically address spontaneously occurring self-talk (Latinjak et al., 2019; Van Raalte et al., 2016) which is more frequent and arguably more relevant to non-elite athletes. Endurance sport is particularly interesting from a cognitive perspective because it is a real-world example of a situation traditionally thought to require a high degree of cognitive control. In the presence of unavoidable fatigue, long-distance runners, cyclists, swimmers, rowers, etc. have to continuously inhibit the prepotent response (slowing down or quitting) in order to fulfil a longer-term goal. These athletes presumably also have rich opportunity for self-talk content as they are often alone with their thoughts for prolonged stretches of time during both training and competition.

There are several unresolved issues with the self-talk intervention studies that warrant further investigations before a causal link between how people talk to themselves and how they perform can be established. First, the studies are often underpowered with only a few participants in each intervention condition. A metaanalysis by Hatzigeorgiadis et al. (2011) suggested that the average effect size of self-talk interventions is a Cohen's *d* of 0.48. With this kind of medium-sized effect and between-subjects design, an example power analysis suggests that a study would need approximately 69 participants in each group to detect a difference between two intervention groups with a power of 0.8[1]. A sample size such as this has been the exception rather than the rule (Schweizer & Furley, 2016). With fewer participants, there is an increased risk of both false positives – finding an effect that is not truly there – and false negatives – neglecting to find an effect which is in fact present (Świątkowski & Dompnier, 2017). Second, the intervention studies also in many cases lack active control groups and simply compared participants who had undergone self-talk training and participants who had not undergone any training. The inclusion of active control groups is important because of potential placebo effects. Due to the design of most of these intervention studies, it has not been possible to conclude that the self-talk

[1] This sample size analysis was conducted for a two-tailed, two-sample t test with alpha = 0.05 using the *pwr* library in R (Champely, 2020).

interventions directly caused performance improvement – it could also simply be the case that undergoing any intervention will help, regardless of the content.

1.3. The present study

We aimed to apply the dual-task interference method from cognitive psychology to the interesting question of how people motivate themselves for physical endurance. While no studies to date have used dual-task interference specifically to test the role of inner speech in endurance performance, there *are* dual-task costs associated with a diverse range of physical performance measures such as jump landing performance (Biese et al., 2019), single-leg postural control (Talarico et al., 2017), climbing (Epling et al., 2018; Green & Helton, 2011; Woodham et al., 2016), swimming (Stets et al., 2020), and running (Blakely et al., 2016). Even though there is evidence of dual-task interference between physical and cognitive tasks, the nature of this interference remains underdetermined – the interference tasks could disrupt mental imagery, inner speech, or attentional mechanisms generally. In the present study, we use a dual-task paradigm specifically designed to investigate the contribution of inner speech to endurance.

In designing the present study, we noted that many of the verbal interference methods used in the literature are not suitable for sports. Articulatory suppression (constantly saying "the" out loud), for example, would introduce a serious confound by interfering with respiration. The simple motor control often used as comparison – foot tapping – would similarly comprise a motor confound. In the first experiment, we therefore used two memory tasks (memory for letters and numbers or memory for locations on a grid). Aside from not interfering with breathing, these interference tasks had the advantage that we were able to assess performance on them to control for trade-off effects (see Nedergaard et al., 2022, for a discussion). However, as can be seen in more detail in section 3.4. Interim Discussion below, the first experiment had some methodological weaknesses, notably that the interference tasks were not continuous. Because of the methodological weaknesses, we conducted a second experiment with verbal and visuospatial 2-back matching tasks as interference tasks. Our preregistered hypotheses for Experiment 1 (https://osf.io/2ah7s; Experiment 2: https://osf.io/byfp3) were as follows:

I. Cycling performance will decrease in both the verbal and non-verbal interference conditions compared to the control condition.

II. If inner speech is required to maximise performance, we expect cycling performance to decrease significantly more in the verbal compared to the non-verbal interference condition.

III. If there is no detectable dual-task effect on cycling performance, we expect to see a trade-off where there is instead a detrimental effect on the verbal or non-verbal simultaneous task.

IV. Participants who indicate high self-talk frequency and efficacy in the questionnaire will be more negatively affected by the verbal distraction task than other participants.

2. EXPERIMENT 1: METHOD

To ensure transparency and accountability, we preregistered this study on the Open Science Framework. We chose to aim for approximately 50 participants as this seemed reasonable given our within-subjects design and the moderate effect sizes found in the verbal interference literature (Brysbaert, 2019; Nedergaard et al., 2022; Schweizer & Furley, 2016). For other interference studies related to physical control, the sizes of the interference effects have been in the $d = 0.3$ to $d = 0.7$ range (Biese et al., 2019; Talarico et al., 2017). Simulated analyses further indicated that 50 participants would be sufficient to detect an effect size of $d = 0.4$ (the estimated 'smallest effect size of interest'; Brysbaert, 2019) (script available on OSF). Repeated measures designs such as ours require fewer participants than the between-groups designs used in the majority of intervention studies.

2.1. Participants

The project received ethical approval from both the Institutional Review Board at Aarhus University and the Human Subjects Committee at the Cognition and Behavior Lab at Aarhus University. We recruited 49 participants from the participant pool attached to Cognition and Behavior Lab. Participants were all above 18 years of age, normally exercised at least twice a week, and reported no known heart conditions (median age = 24 y; range = 18 to 76 y; 29 men and 20 women). Especially the exercise requirement constrains the generalisability of our results as there may be different relationships between inner speech and physical performance for people who do not exercise regularly. However, we chose to implement this requirement to avoid unnecessary risk to participants. Given the wide age range, relative gender balance, and variety of nationalities (31 Danish and 18 non-Danish), we believe our results are relatively generalisable. Participants received 90 DKK as compensation for their time. Participants were asked to measure their resting heart rate prior to the experiment. Nine participants had not measured this, so it was estimated based on their age, gender, and exercise frequency (see Reimers, Knapp, & Reimers, 2018; Quer et al., 2020).

2.2. Materials

Transparency and openness. All data and PsychoPy code for the experiment can be accessed at the Open Science Framework (https://osf.io/uk2y4/?view_only=6fc8f12830df497e9c403cfb01ebc66c). The data for Experiment 1 were collected in 2020.

Cycling. We ran the experiment using custom-written software in PsychoPy version 3.2.4 (Peirce, 2007). The exercise bike was a Titan Fitness model SB550 Prestige adjusted to Level 14 resistance (piloting had shown that this level of resistance suited the widest range of participants). We used a CatEye Velo 7 cycling computer (CatEye, Osaka, Japan) attached to the exercise bike to measure meters per trial.

Heart rate. We used a Charge 2 FitBit (Fitbit, San Francisco, California, USA) wristband to measure heart rate during the experiment. While wrist-worn heart rate monitors are not as accurate as chest-worn monitors, we opted for the wrist-worn monitor for convenience – we did not need high-fidelity accuracy but simply to have a way of making sure that participants were putting in effort. Benedetto et al. (2018) tested the accuracy of the FitBit Charge 2 wristband and found that it had a modest bias in measuring heart rate at -5.9 bpm (95%CI: -6.1 to -5.6 bpm). We therefore added 5.9 bpm to all heart rate measures. We were unable to retrieve heart rate data from eight participants so their heart rate data were excluded from subsequent analyses. All participants were instructed to reach 70 % of their heart rate reserve on each cycling trial. We calculated 70 % of the individual participant's heart rate reserve with the following formula (adapted from Tanaka, Monahan, & Seals, 2001):

$$HR_{target} = ((208 - 0.7 * age) - HR_{rest}) * 0.7 + HR_{rest}$$

Questionnaire. Participants completed the Automatic Self-talk Use Questionnaire for Sports ASTQS (Zourbanos et al., 2009) prior to the cycling section of the experiment. The ASTQS is a questionnaire made to measure the quantity and quality of self-talk used by athletes of varying levels of activity and fitness. The questionnaire measures four positive and four negative self-talk dimensions. Positive self-talk consists of psych-up (e.g., 'come on'), confidence (e.g., 'I'm very well prepared'), anxiety-control (e.g., 'don't get upset'), and instruction (e.g., 'concentrate on what you have to do right now' while negative self-talk consists of worry (e.g., 'I'm going to lose'), disengagement (e.g., 'I can't keep going'), somatic fatigue (e.g., 'I'm tired'), and irrelevant thoughts (e.g., 'what am I doing later today?').

Analysis. All analyses were conducted in R version 4.1.3 (R Core Team, 2022) and RStudio version 2022.02.3. All plots were drawn with *ggplot2* (Wickham, 2016) and all linear models were constructed with *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017).

2.3. Procedure

Participants began the experiment by filling out the ASTQS and then proceeded to the cycling section of the experiment. After a brief warm-up and an introduction to the experimental set-up, participants completed 24 1-minute trials (12 rest and 12 cycling, interleaved). Previous studies have indicated that a 1 minute-sprint is a sufficient duration to require endurance control (Craig et al., 1989; Martin et al., 2007). See also Figure 1 for a sketch. During each 1-minute trial, participants were asked to rehearse and remember either the locations of six letters and numbers on a grid (visuospatial) or the letters and numbers themselves (verbal). A third of the trials were control trials where participants did not have to remember anything. The stimuli were presented in the same way regardless of the verbal or visuospatial nature of the memory task: Six letters and numbers were randomly selected by the computer and appeared sequentially for one second each. After the stimuli were presented, the program counted down from three and started a 1-minute countdown on the computer screen for the duration of the trial. When the countdown had finished, participants had as much time as they wanted to click on either the locations they remembered (visuospatial trial) or the letters and numbers they remembered (verbal trial). When responding after verbal interference trials, the letters and numbers appeared in new locations that were unrelated to the locations in which they were originally presented.

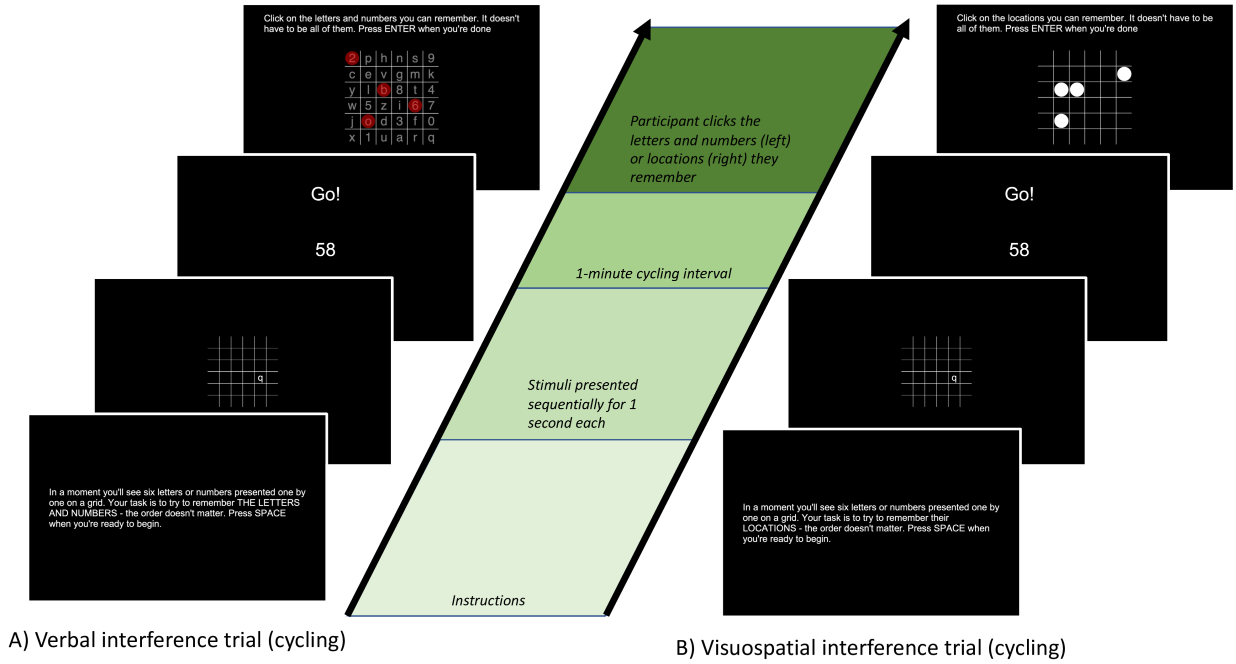


Figure 1. *Schematic of the procedure in Experiment 1. Figure 1A on the left shows a cycling trial with verbal interference while Figure 1B on the right shows a cycling trial with visuospatial interference.*

3. EXPERIMENT 1: RESULTS

3.1. Descriptive statistics

Questionnaire. Only three of our 49 participants answered that they never talk to themselves while exercising. Of the remaining 46, six answered that they 'rarely' talk to themselves while exercising, 24 said that they 'sometimes' talk to themselves while exercising, 12 said that they 'often' do so, and four said that they 'very often' do so. In terms of self-talk efficacy, 19 participants reported that self-talk usually has a positive effect on their performance, 21 said that the effect is sometimes positive and sometimes negative, and seven said that self-talk does not affect their performance. See Figure 2 for a visualisation of self-talk frequency and experienced efficacy.

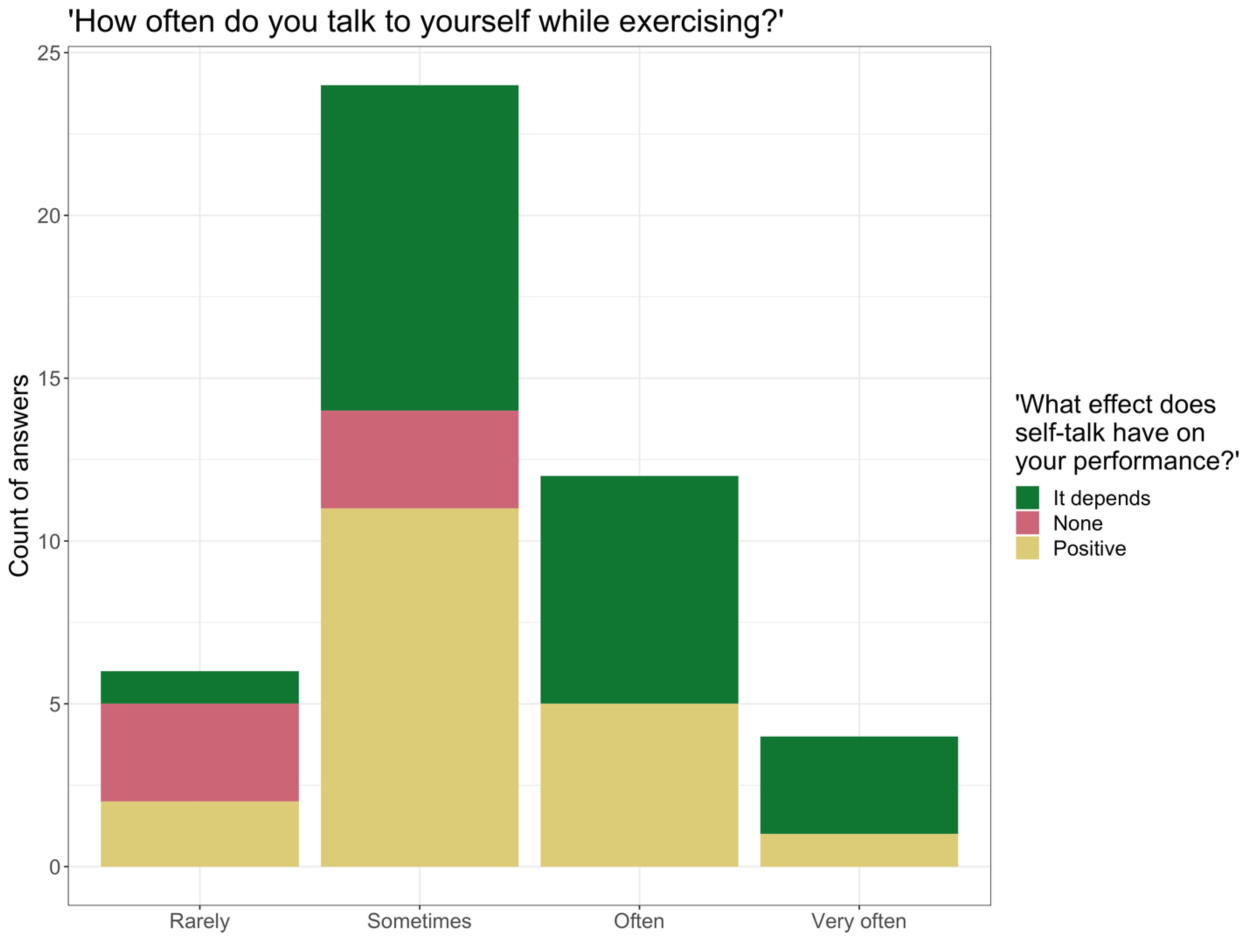


Figure 2. *Bar plot of participants' answers to the self-talk efficacy and self-talk frequency questionnaire items.*

For exercise frequency, one participant reported exercising a few times a month, 15 participants reported exercising a few times a week, 21 participants reported exercising most days every week, 12 participants reported exercising every day or almost every day, and one participant reported exercising several times almost every day. There were no significant differences within either cycling or memory performance depending on the level of exercise frequency, as indicated by linear mixed models of exercise predicting meters cycled (p = 1) and memory performance (p = .971).

Heart rate. The heart rate data was low-pass filtered using a Butterworth filter with an order of 5 and a cut-off frequency of 0.05 Hz (20s) using he 'filtfilt' and 'butter' functions from R package *gsignal* (Van Boxtel & et al., 2021). We used the 'findpeaks' function from the R package *pracma* to determine both peaks and troughs in heart rate (Borchers, 2021). Out of a total of 473 valid cycling trials (see above), participants reached the target of 70 % maximal heart rate on 326 trials (68.9 %) and did not reach the target on 147 trials. An independent samples t-tests indicated no difference between trials where the target was reached and where the target was not reached for memory performance ($t(198.12) = -1.365$, $p = .174$). A chi-squared test also confirmed that there was no difference between interference conditions in terms of the proportion of trials on which the target was reached ($\chi^2(2) = 0.26$, $p = .876$). As is evident from Figure 3 below, there was a large difference (> 2 SDs) between heart rate peaks during cycling and heart rate troughs during rest. Given the very short restoration time (less than one minute), we can therefore be confident that participants did indeed put sufficient pressure on themselves during cycling trials to demand a certain degree of executive control. For each of the subsequently reported tests, we also tested whether the effects were different between trials where the target was reached and where it was not – this was never the case.

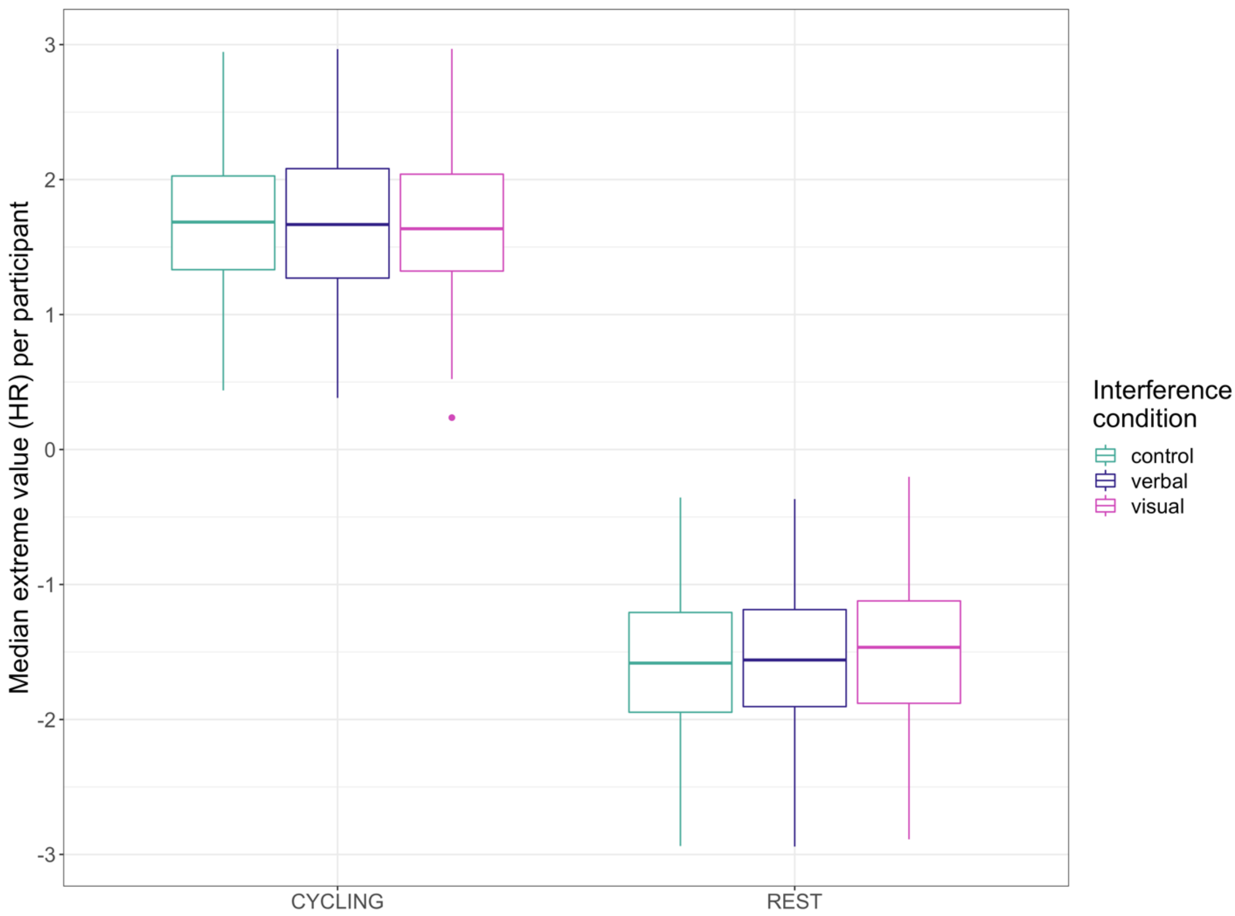


Figure 3. *Boxplot showing z-scored heart rate during cycling versus rest in the three interference conditions. The upper and lower hinges correspond to the first and third quartiles, and the central tendency line indicates the median. The upper and lower whiskers extend to a distance of 1.5 * the inter-quartile range (the middle half of the distribution).*

Interference tasks. Participants performed better on the verbal interference task than on the visuospatial interference task. See Table 1 for an overview of participants' performance on the memory tasks during cycling intervals and rest intervals and Figure 4 for a visualisation of the same. To test whether participants' performance was above chance, we simulated 100000 trials of six "clicks" with a $\frac{6}{36}$ probability of each click being correct. This probability is higher than it should be as participants in the actual experiment sampled without replacement but this is to allow for the fact that participants could change their mind about their responses. Through this procedure, we established that participants should get one correct click on average each trial if they picked six randomly (average success = 0.17). A trial was significantly above chance ($p < .05$) if it had 3 or more correct clicks. A Wilcoxon rank sum test of the difference between the simulated means and memory performance from the experiment showed that performance was significantly above chance on both the visuospatial interference task ($W = 5156398$, $p < .001$) and the verbal interference task ($W = 954200$, $p < .001$). We conducted this non-parametric test as the data were not normally distributed.

Table 1. *Performance on the interference tasks during cycling and rest.*

Interference condition	Cycling condition	Mean % success	Median % success	SD of % success
verbal	REST	0.86	1	0.21
verbal	CYCLING	0.84	1	0.24
visual	REST	0.54	0.5	0.28
visual	CYCLING	0.52	0.5	0.27

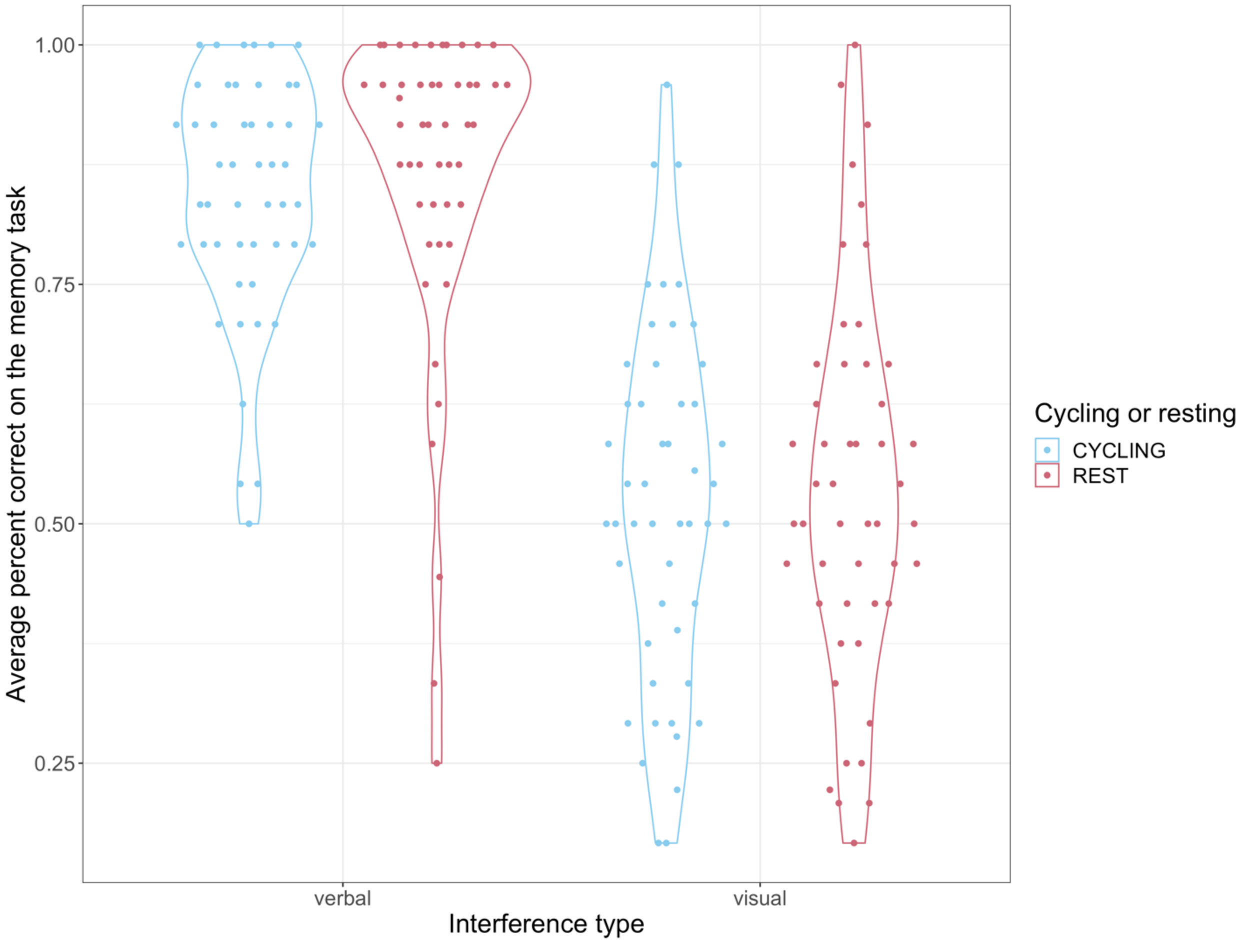


Figure 4. *Violin and scatter plot showing participants' performance on the two interference tasks during cycling and rest trials.*

Cycling performance. Participants generally cycled furthest in the control condition (M = 214.77 meters) followed by the visuospatial interference condition (M = 213.31 meters) and the verbal interference condition (M = 212.60 meters). See also Figure 4. We scaled the meters cycled according to the individual participant's mean distance cycled to control for individual fitness levels. These scaled meters are used in subsequent analyses and models, both because it allowed us to control for differences in fitness levels and because the scaled meters met normality assumptions and the untransformed meters cycled did not. This transformation was not included in the preregistration for Experiment 1.

We also calculated individual susceptibility to verbal interference by subtracting within-person average performance on verbal interference trials from within-person average performance on visual interference trials. This was used to compare to participants own experience of effects of inner speech on performance.

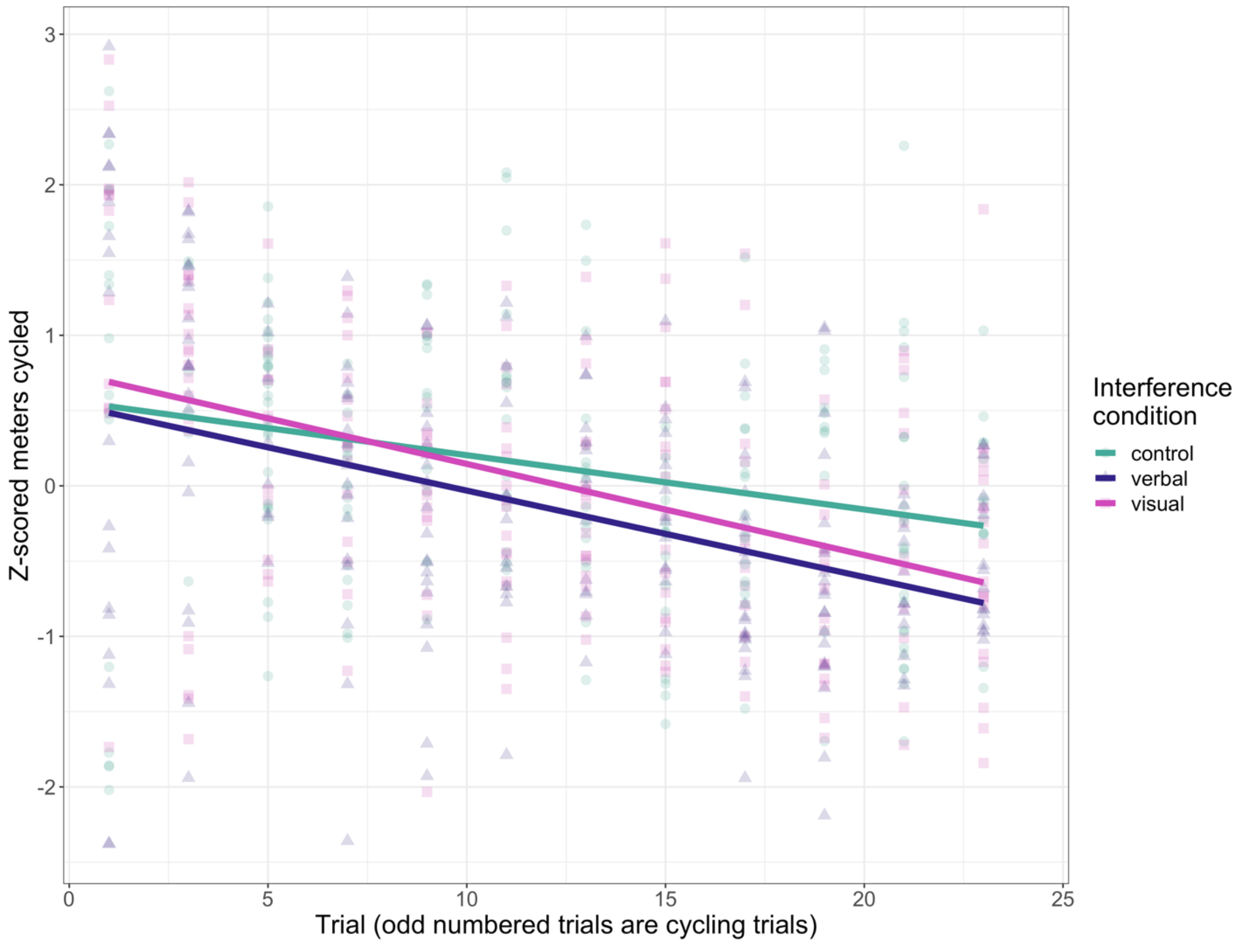


Figure 5. *Plot showing participants' z-scored cycling performance across the entire experiment (12 cycling trials). The three lines represent linear models of performance during verbal interference, visuospatial interference, and a no-interference control condition. Points indicate individual performance on a given trial.*

3.2. Preregistered linear mixed models

Dual-task condition predicting cycling performance. We conducted a linear mixed model of dual-task condition predicting z-scored meters cycled including random intercepts for participants and random slopes for trial. This model suggested that the participants in the control condition cycled significantly faster than in the verbal interference condition (β = 0.27; *SE* = 0.10; *t*(432.51) = 2.85; *p* < .001; see also Figure 5). There was no significant difference between either the visuospatial interference condition and the verbal interference condition (*p* = .10) or between visuospatial interference and the control condition (*p* = .227). Cohen's *d* for the difference between verbal interference and control trials was 0.29 while Cohen's *d* for the difference between verbal and visual interference trials was 0.22. We calculated effect sizes using the 'cohen.d' function from the *effsize* package in R (Torchiano, 2020).

Self-reported self-talk frequency and self-talk efficacy predicting verbal interference. We calculated degree of verbal interference by subtracting mean meters cycled in the verbal interference condition from mean meters cycled in the visual interference condition for each participant. Thus, a positive coefficient indicates that verbal interference was more detrimental than visual interference, and a larger difference suggests a stronger effect of verbal interference for the individual. The linear model with self-talk frequency (treated as a numeric predictor) predicting degree of interference found no effect of increased self-talk frequency on degree of interference (*p* > .361). See Figure 6A.

For the linear model of self-talk efficacy (treated as a categorical predictor) predicting degree of interference, there were again no differences between groups ('None' versus 'It depends': p = .085; 'Positive' versus 'It depends': p = .227). See Figure 6B for an illustration of differences.

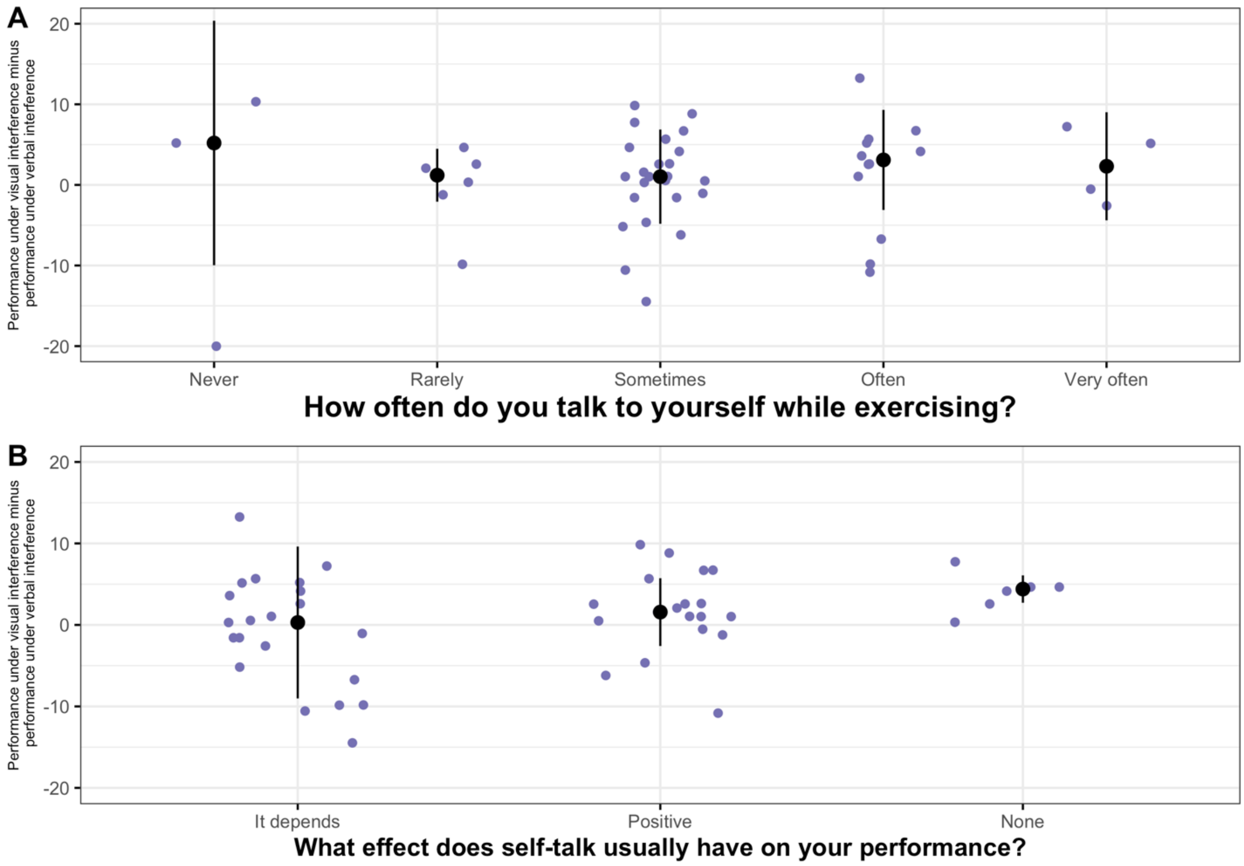


Figure 6. *Line and jitter plots showing the difference between susceptibility to verbal interference (meters cycled under visual interference minus meters cycled under verbal interference) as a function of the frequency (A) and efficacy (B) of self-talk. Dots indicate median, error bars indicate interquartile range.*

3.3. Trade-off between interference task and cycling performance

To ascertain whether there was a trade-off between the interference tasks and cycling performance, we conducted linear mixed model with z-scored meters cycled and interference condition predicting z-scored accuracy on the interference tasks. This model included random slopes over trials per participant. There was evidence that participants performed less well in the visual interference condition compared to the verbal interference condition (β = -1.07, *SE* = 0.08, t(333.16) = -13.22, p < .001). However, there was no effect of z-scored meters cycled on interference task performance (p = .230) and no significant interaction between interference condition and z-scored meters cycled (p = .573). See Figure 7.

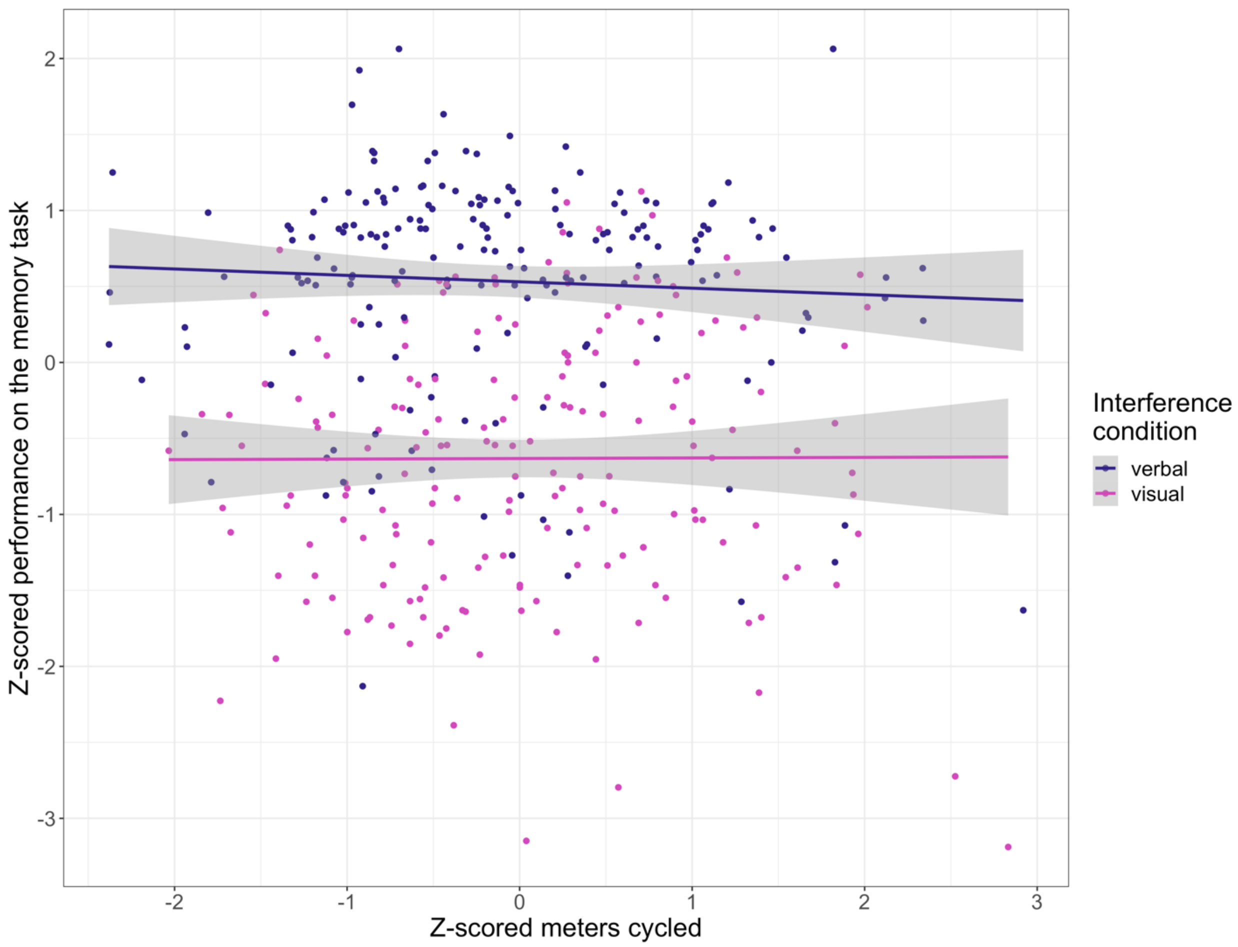


Figure 7. *Scatterplot showing the correlation between meters cycled (scaled according to individual participant) and performance on the verbal and visuospatial memory tasks (also scaled according to individual participant). No signs of systematic trade-offs were found. Shaded areas indicate 95 % confidence intervals.*

3.4. Interim discussion

As hypothesised, we found that verbal interference had a detrimental effect on cycling performance. This effect, however, was only statistically significant when comparing against the no interference condition and not against the visuospatial condition. There may be different reasons for this. The first option to consider, of course, is that our hypothesis about the involvement of inner speech in physical exercise is wrong. However, given previous findings reviewed in the introduction and the fact that we find a nominal effect pointing in the right direction, we are reluctant to accept this without further considerations. Unfortunately, we also observed a difference in task difficulty for the two interference tasks, resulting in ceiling effects for performance on the verbal interference task, which were not found for the visuospatial interference task (see Table 1). The confounding difference in attentional demand between the two tasks could have led to an underestimation of the effect of verbal interference relative to visual interference. The visual presentation of the stimuli in the beginning of the trial may also have enabled a non-verbal storage strategy that did not involve the articulatory system, thus not interfering as strongly with inner speech as expected. Alternatively, participants might have been able to use some sort of long-term storage which also allowed them to continue their use of their inner voice during the experiment, to some degree. Lastly, the visuospatial task might have been so difficult that participants down-prioritised the visual memory task during the trial because it was too difficult. The latter, however, is not supported by data. Performance on the visual memory task was well above chance level as established through simulations, and second, there was no evidence for a trade-off between interference task and cycling.

If the effect of verbal interference found in this experiment is real and robust, then a more continuous interference task should cause a larger effect of verbal interference. Thus, we decided to conduct a follow-up experiment with interference tasks with continuous interference during the cycling trial. This yields the added benefit of allowing us to make conceptual comparisons between effects of different kinds of interference tasks which has rarely been done in previous research (Bek et al., 2009, 2013; Nedergaard et al., 2022; Piccardi et al., 2020; Roberson & Davidoff, 2000). In Experiment 2, we also measured ECG with electrodes to get more accurate physiological measures than those obtained from the FitBit2 wristband and used a cadence sensor attached to the bike to get more fine-grained performance data.

4. EXPERIMENT 2: METHOD

We once again preregistered this study on the Open Science Framework (https://osf.io/byfp3). Our hypotheses were the same as for Experiment 1. We aimed for approximately 50 participants again for the same reasons as detailed in the Method section for Experiment 1, and because we hypothesised that continuous interference would yield a stronger effect in the verbal versus visuospatial interference contrast.

4.1. Participants

The project received ethical approval from both the Institutional Review Board at Aarhus University and the Human Subjects Committee at the Cognition and Behavior Lab at Aarhus University. We recruited 50 participants from the participant pool attached to Cognition and Behavior Lab. Participants were all above 18 years of age, normally exercised at least twice a week, and had no known heart conditions (median age = 25 y, range = 18 to 36 y; 32 men, 17 women, and one who preferred not to disclose their gender). Given the relative gender balance, and variety of nationalities (30 Danish and 20 non-Danish), we believe our results are relatively generalisable. Participants received 110 DKK as compensation for their time (more than in Experiment 1 because the improved physiological measures took longer to set up). Ten participants had not measured their resting heart rate prior to the experiment so it was estimated based on their age, gender, and exercise frequency (see Reimers, Knapp, & Reimers, 2018; Quer et al., 2020).

4.2. Materials

Transparency and openness. All data and PsychoPy code for the experiment can be accessed at the Open Science Framework (https://osf.io/uk2y4/?view_only=6fc8f12830df497e9c403cfb01ebc66c). The data for Experiment 2 were collected in 2022.

Cycling. We ran the experiment using custom-written software in PsychoPy version 3.2.4. The exercise bike was a Titan Fitness model SB550 Prestige adjusted to Level 14 resistance (identical to Experiment 1). We used a Wahoo RPM Cadence Censor v1.54.0.10 (Wahoo Fitness, Atlanta, Georgia, USA) attached to the exercise bike to measure cadence.

Heart rate. We used a BIOPAC BioNomadix system (BIOPAC Systems, Goleta, California, USA) to measure heart rate and respiration during Experiment 2. All participants were instructed to reach 70 % of their heart rate reserve on each cycling trial. We calculated 70 % of the individual participant's heart rate reserve with the following formula (Tanaka et al., 2001):

$$HR_{target} = ((208 - 0.7 * age) - HR_{rest}) * 0.7 + HR_{rest}$$

Questionnaire. Participants completed the Automatic Self-talk Use Questionnaire for Sports (Zourbanos et al., 2009) prior to the cycling section of the experiment. See Materials section for Experiment 1 above for a description of this questionnaire.

Analysis. All analyses were conducted in R version 4.1.3 (R Core Team, 2022) and RStudio version 2022.02.3. All plots were drawn with *ggplot2* (Wickham, 2016) and all linear models were constructed with *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017).

4.3. Procedure

Participants began the experiment by filling out the ASTQS and then proceeded to the cycling section of the experiment. After a brief warm-up and an introduction to the experiment set-up, participants completed 24 1-minute trials (12 rest and 12 cycling, interleaved). See also Figure 8 for a sketch. During each 1-minute interference trial, participants performed a 2-back memory task. They were asked to pay attention to either nonsense words played every other second (verbal interference) or coloured, geometric figures appearing in different locations on the screen every other second (visuospatial interference). Participants had to press a button (attached to the handles of the stationary bike) if the word they heard or the figure they saw was the same as the one two before. A third of the trials were control trials where participants did not have to remember anything. They were, however, required to press the button every 10 seconds to control for motor interference. Instead of a count-down of the seconds presented on the screen with numbers as in Experiment 1, Experiment 2 counted down the trial duration with a blue bar at the top of the screen. The cue to whether participants were in a cycling trial or a resting trial was also non-verbal in Experiment 2 contrasting with Experiment 1.

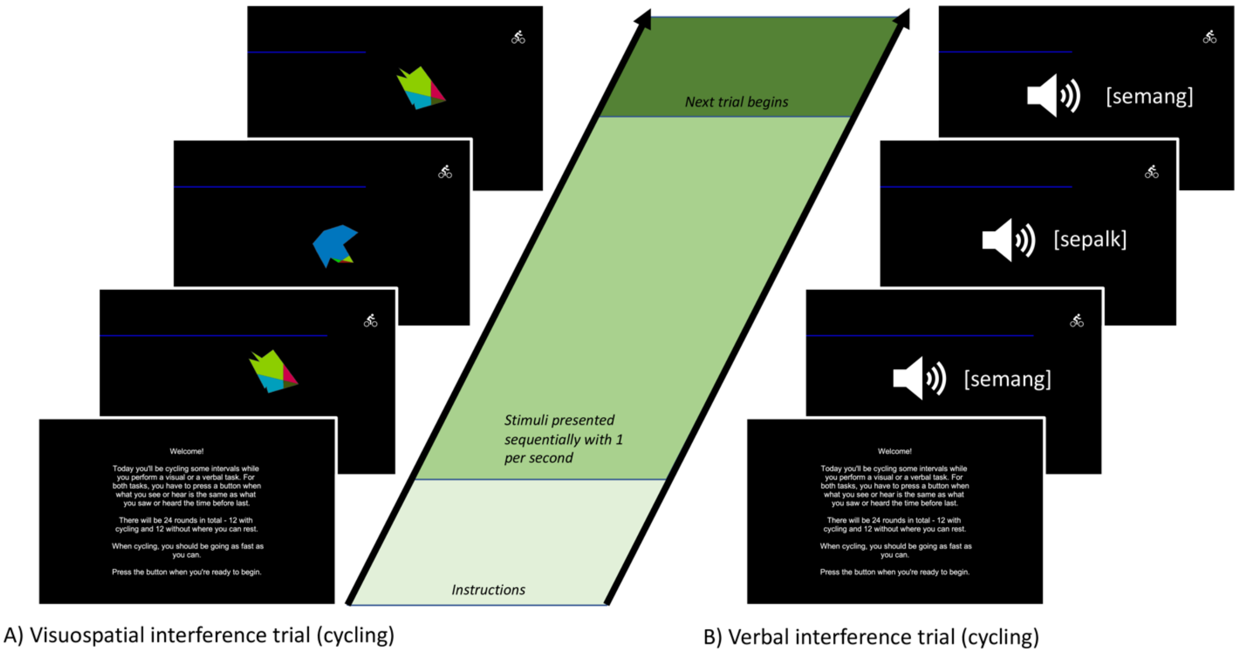


Figure 8. *Schematic of the procedure in Experiment 2. Figure 8A on the left shows a cycling trial with visuospatial interference while Figure 8B on the right shows a cycling trial with verbal interference. Both show examples of 2-back matches. Note that the 2-back matching nonsense words were only presented auditorily in the actual experiment.*

5. EXPERIMENT 2: RESULTS

5.1. Descriptive statistics

Questionnaire. Only four of our 50 participants answered that they never talk to themselves while exercising. Of the remaining 46, seven answered that they ‘rarely’ talk to themselves while exercising, 10 said that they ‘rarely’ talk to themselves while exercising, 21 said that they ‘sometimes’ do so, nine said that they ‘often’ do so, and six said that they ‘very often’ do so. In terms of self-talk efficacy, 17 participants reported that self-talk usually has a positive effect on their performance, 19 said that the effect is sometimes positive and sometimes negative, six said that self-talk does not affect their performance, and four said that self-talk usually has a negative effect on their performance. See Figure 9 for a visualisation of self-talk frequency and experienced efficacy.

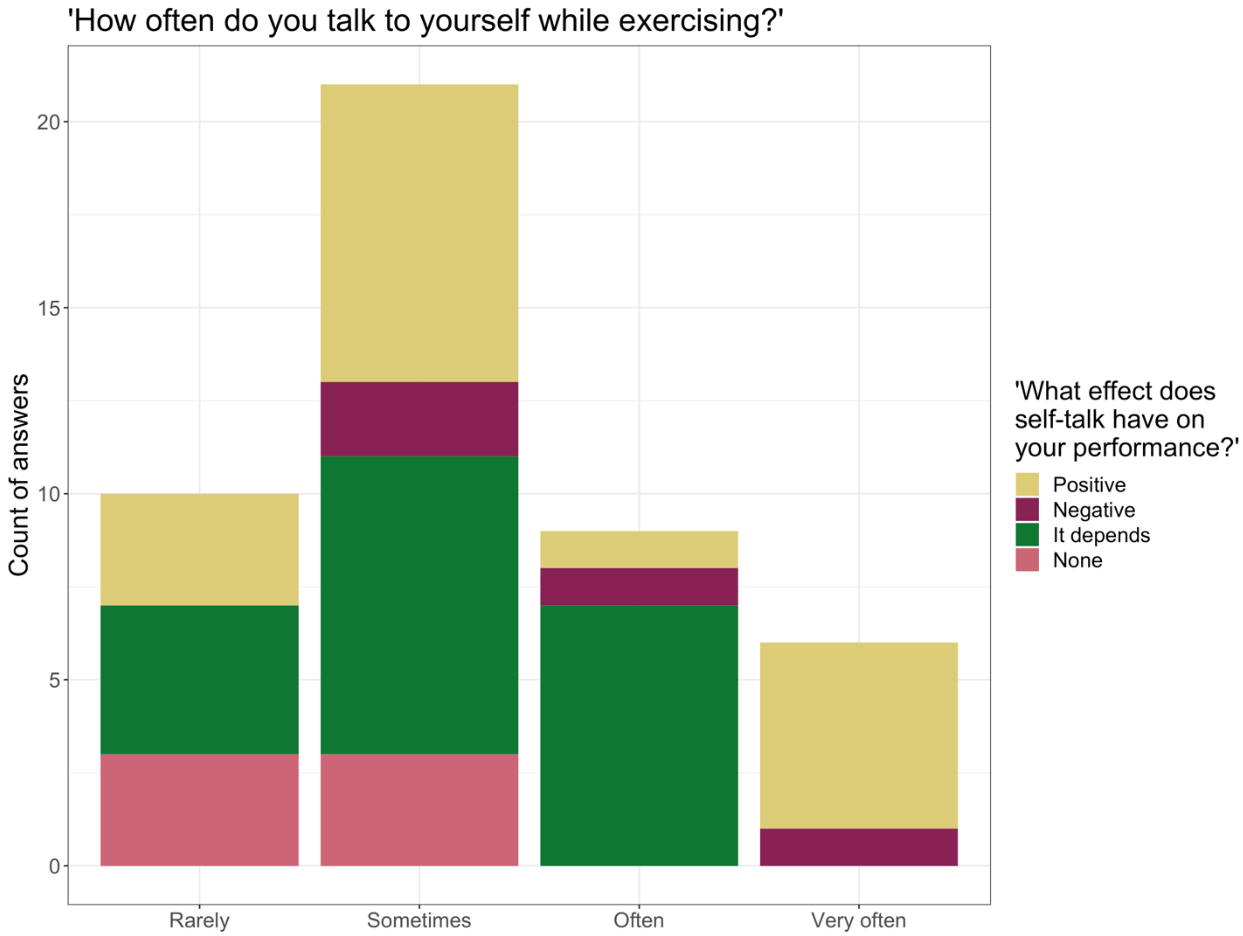


Figure 9. *Visualisation of participants' answers to the self-talk efficacy and self-talk frequency questionnaire items in Experiment 2.*

Heart rate. The heart rate data was low-pass filtered using a Butterworth filter with an order of 5 and a cut-off frequency of 0.05 Hz (20s) using he 'filtfilt' and 'butter' functions from R package *gsignal* (Van Boxtel & et al., 2021). We used the 'findpeaks' function from the R package *pracma* to determine both peaks and troughs in heart rate (Borchers, 2021). Due to technical difficulties, we excluded heart rate data from one participant. Out of a total of 586 valid cycling trials, participants reached the target of 70 % maximal heart rate on 444 trials (75.8 %) and did not reach the target on 142 trials. An independent samples t-tests indicated no difference between trials where the target was reached and where the target was not reached for *d'* memory performance ($t(129.56) = 0.74$, $p = .459$). A chi-squared test also confirmed that there was no difference between interference conditions in terms of the proportion of trials on which the target was reached ($\chi^2(2) = 0.13$, $p = .938$). As is evident from Figure 10 below, there was a large difference (> 2 SDs) between heart rate peaks during cycling and heart rate troughs during rest. Given the very short restoration time (less than one minute), we can therefore be confident that participants did indeed put sufficient pressure on themselves during cycling trials to demand a certain degree of executive control. For each of the subsequently reported tests, we also tested whether the effects were different between trials where the target was reached and where it was not – this was never the case.

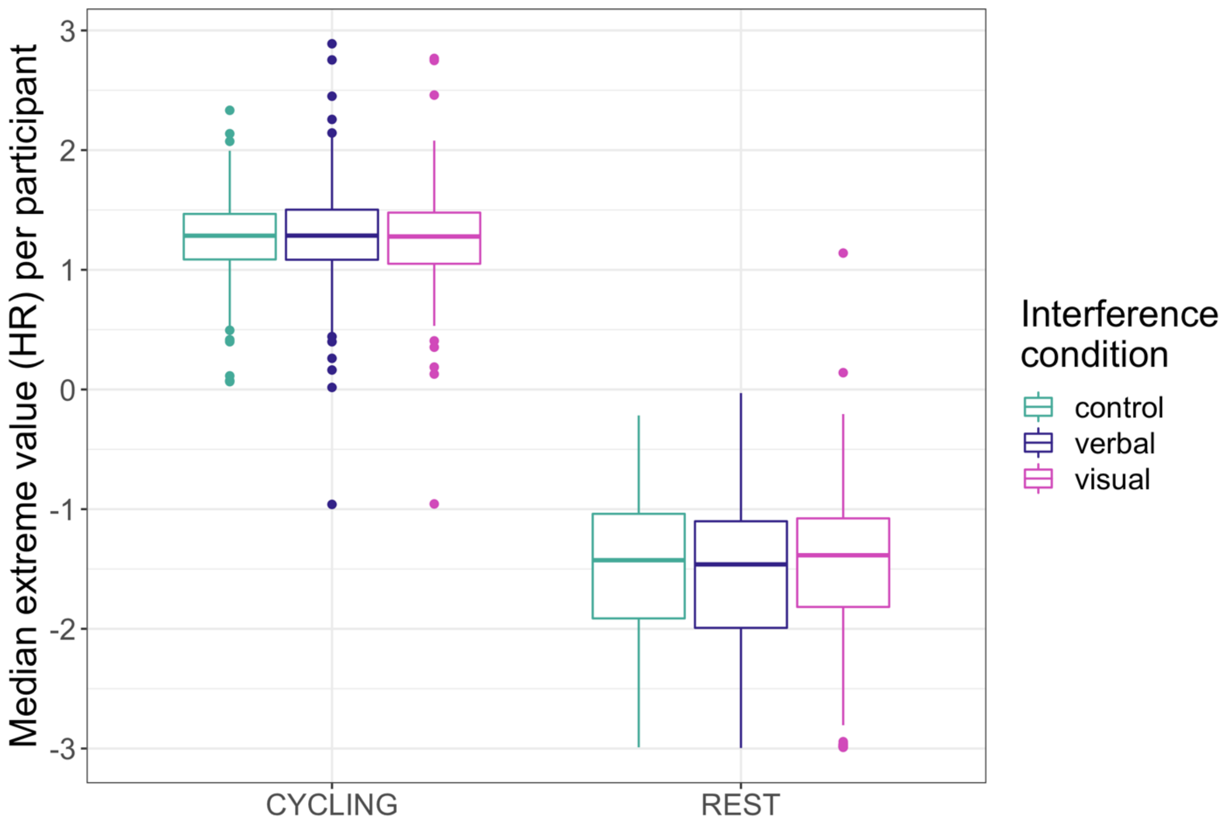


Figure 10. *Boxplot showing z-scored heart rate during cycling versus rest in the three interference conditions. The upper and lower hinges correspond to the first and third quartiles, and the central tendency line indicates the*

*median. The upper and lower whiskers extend to a distance of 1.5 * the inter-quartile range (the middle half of the distribution).*

Interference tasks. Participants performed better on the verbal interference task than on the visuospatial interference task. See Table 2 for an overview of participants' performance on the memory tasks during cycling intervals and rest intervals and Figure 11 for a visualisation of the same. On many individual trials, participants had 100 % hit rate which creates infinite *d'* estimates. To prevent this, we used the adjustment (Hautus, 1995; Stanislaw & Todorov, 1999) built into the 'dprime' function from the *psycho* package in R (Makowski, 2018).

Table 2. *Performance on the interference tasks during cycling and rest.*

Interference condition	Cycling condition	Mean hits out of 6	Mean false alarms out of 24	Mean d'
verbal	REST	4.54	0.84	2.41
verbal	CYCLING	4.56	1.34	2.28
visual	REST	3.85	1.13	2.02
visual	CYCLING	4.06	1.40	2.06

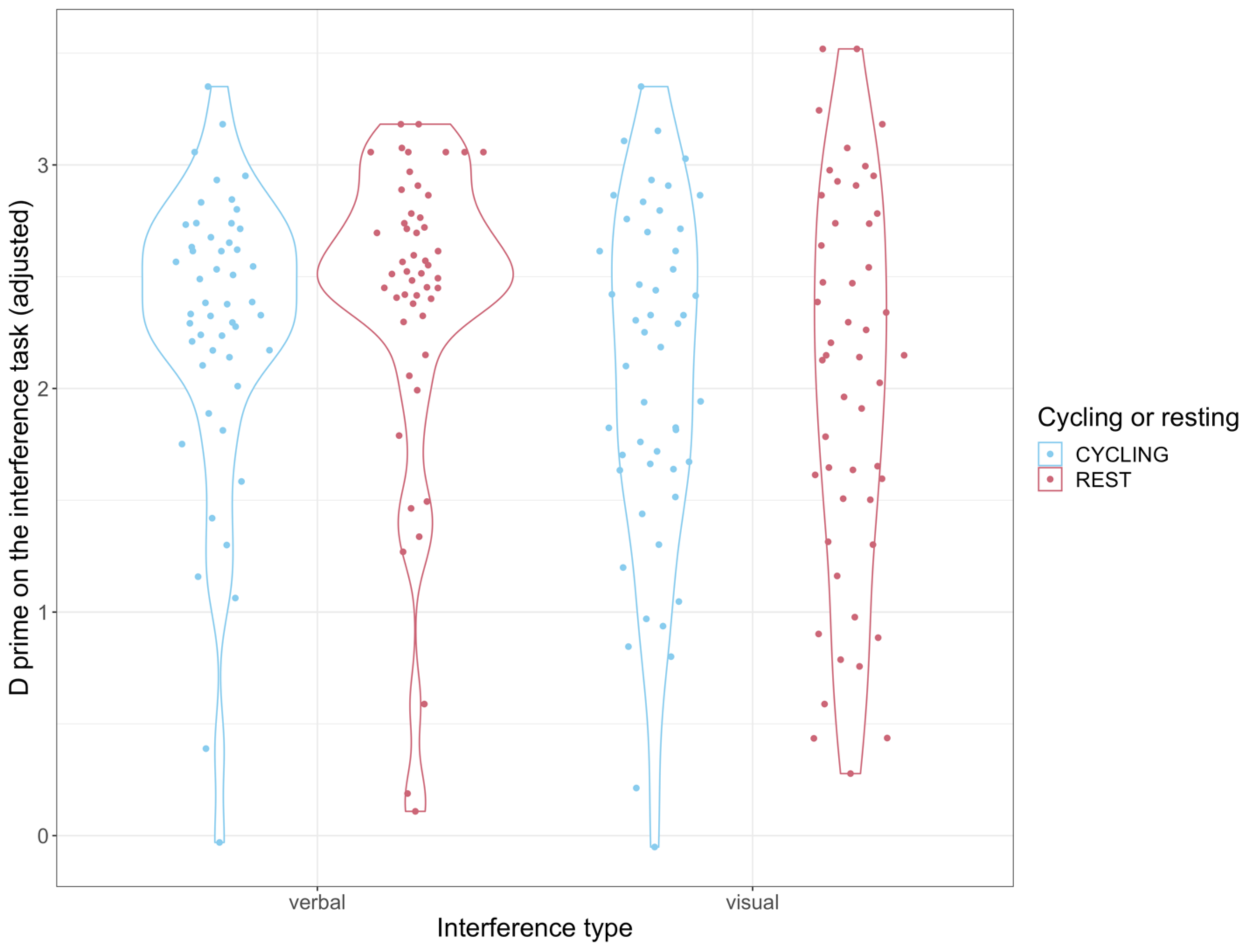


Figure 11. *Violin and jitter plot showing participants' performance* (d') *on the two interference tasks during cycling and break trials. Values of* d' *are adjusted to prevent infinite values (see main text). Absolute perfect performance across all trials would equal a* d' *of 4.5.*

Cycling performance. Participants cycled fastest in the control condition (M = 99.9 revolutions per minute) followed by the visuospatial interference condition (M = 96.6 revolutions per minute) and the verbal interference condition (M = 93.9 revolutions per minute). See also Figure 12. As in Experiment 1, we scaled the cycling performance according to the individual participant to control for individual fitness levels (see also preregistration).

Also following Experiment 1, we additionally calculated individual susceptibility to verbal interference by subtracting within-person average performance on verbal interference trials from within-person average performance on visual interference trials. This was used to compare to participants' own experience of effects of inner speech on performance.

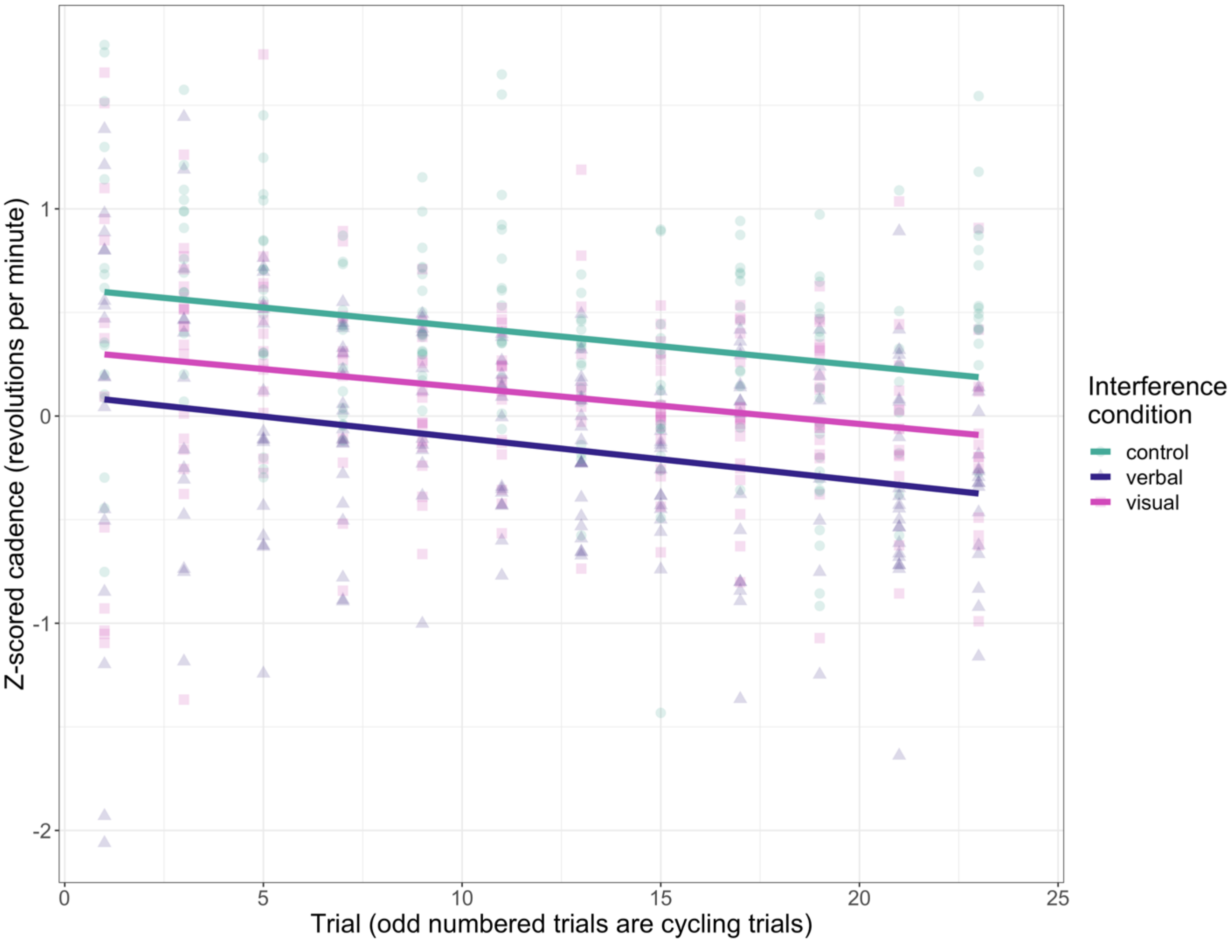


Figure 12. *Plot showing participants' cycling performance across the entire experiment (12 cycling trials), scaled by their individual mean cadence. The three lines represent performance during verbal interference, visuospatial interference, and a no-interference control condition. Points indicate individual performance on a given trial.*

5.2. Preregistered linear mixed models

Dual-task condition predicting cycling performance. Our linear mixed model with scaled cycling cadence (equivalent to cycled meters) as dependent variable and condition as independent variable, including random slopes for trial by participant revealed that the participants in the verbal interference condition cycled with significantly lower cadence than in the control interference condition ($\beta = 0.54$; $SE = 0.05$; $t(501.98) = 10.26$; $p < .001$) and the visuospatial interference condition ($\beta = 0.25$; $SE = 0.05$; $t(501.84) = 4.72$; $p < .001$). Note that the coefficients are positive because the verbal interference condition was treated as the baseline condition. Cohen's *d* for the difference between verbal interference and control trials was 1.00

while Cohen's *d* for the difference between verbal and visual interference trials was 0.43. We calculated effect sizes using the 'cohen.d' function from the *effsize* package in R (Torchiano, 2020).

Self-reported self-talk frequency and self-talk efficacy predicting verbal interference. We conducted a linear model of self-talk frequency (treated as a numeric predictor) predicting degree of interference which found no evidence of a significant relationship (p = .480). For the linear model of self-talk efficacy (treated as a categorical predictor) predicting degree of verbal interference, we again found no evidence of significant differences between groups ('It depends' versus 'Negative': p = .272; 'It depends' versus 'None': p = .742; 'It depends' versus 'Positive': p = .482). See Figure 13 for self-talk frequency responses, self-talk efficacy responses, and their relationships with degree of interference.

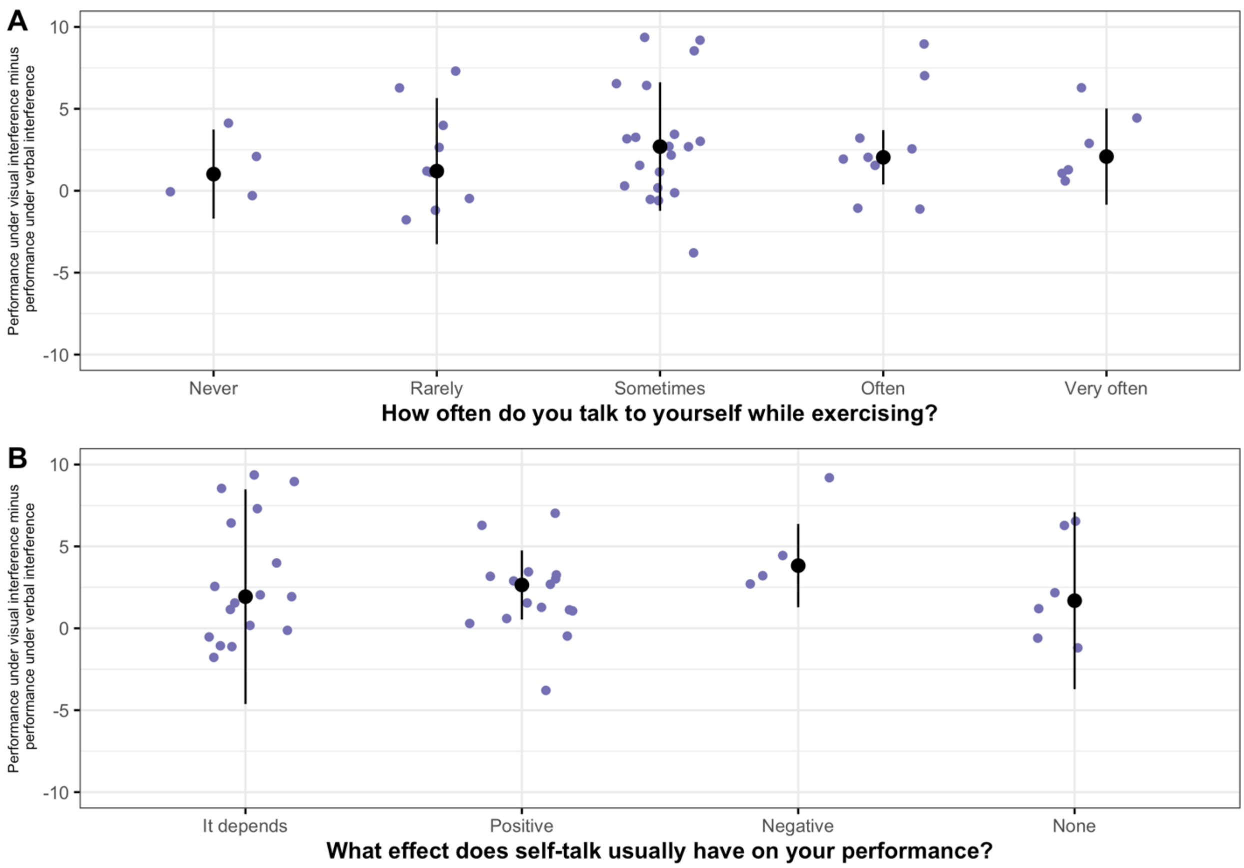


Figure 13. *Line and jitter plots showing the difference between susceptibility to verbal interference (cycling cadence under visual interference minus cycling cadence under verbal interference) as a function of the self-reported frequency (A) and efficacy (B) of self-talk. Dots indicate median, error bars indicate interquartile range. A higher interference score indicates the participants was more negatively affected by verbal interference.*

5.3. Trade-off between cycling performance and 2-back matching performance

To ascertain whether there was a trade-off between the interference tasks and cycling performance, we conducted linear mixed model with z-scored cadence and interference condition predicting *d'* on the interference tasks. This model included a random slope over trials per participant as well as random intercepts for each participant. There was evidence that participants performed less well on the interference task if they cycled faster (β = -0.22, *SE* = 0.10; *t*(296.86) = -2.283, *p* = .023), and participants also performed less well in the visual interference condition compared to the verbal interference condition (β = -0.17, *SE* = 0.07, *t*(298.53) = -2.56, *p* = .011). However, there was no significant interaction between interference condition and z-scored cadence (*p* = .368). See Figure 14. This indicates that the two interference tasks were equally susceptible to trade-off.

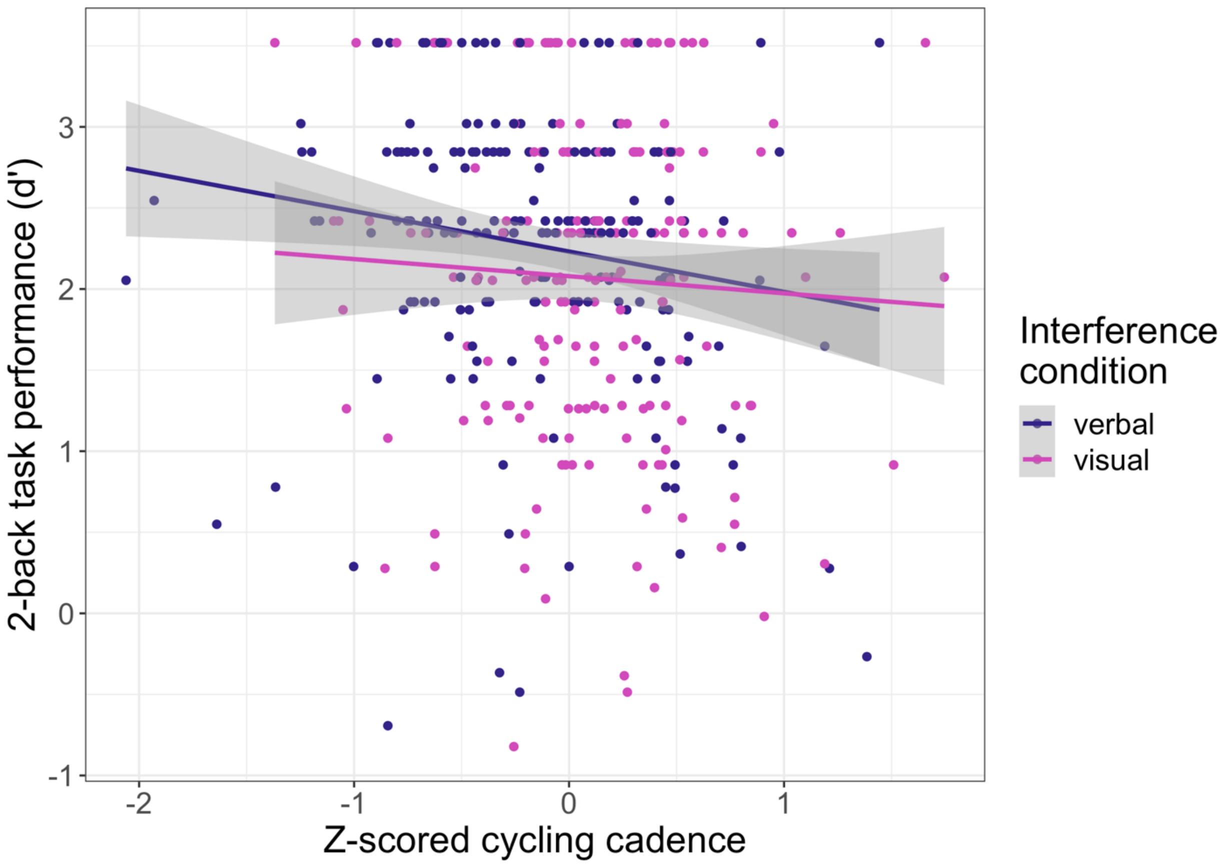


Figure 14. *Scatterplot showing the correlation between meters cycled (scaled according to individual participant) and performance on the verbal and visuospatial memory tasks* (d'*). Shaded areas indicate 95 % confidence intervals.*

6. DISCUSSION

Across two experiments, we found a general effect of cognitive interference on physical endurance performance as well as a specific effect of verbal interference suggesting an important role for inner speech. We tested the influence of four different interference tasks on cycling performance (a "one off" visuospatial memory task, a "one off" verbal memory task, a continuous verbal 2-back matching task, and a continuous visual 2-back matching task). In Experiment 1, which used one-off memory-based interference, only verbal interference had a significant detrimental effect compared with the no-interference control condition (d = 0.29). This effect was nominally larger than the visuospatial effect, but the verbal interference effect (d = 0.22) was not significantly different from the effect of visuospatial interference with the used sample size. Thus, we did not find evidence of a specific role of inner speech. In Experiment 2, which used a continuous interference task with fewer possibilities for adopting non-verbal task strategies, the detrimental effect of verbal interference on cycling performance was stronger than the visuospatial interference (d = 0.43). These results are in line with our main hypothesis. In neither experiment was there an effect of whether participants report that talking to themselves while exercising usually helps them or not.

6.1. Dual-task interference and cognitive control

As discussed in the Introduction, covert language may be involved in endurance performance as a vehicle for behavioural self-cuing, inhibitive control, and motivation. For example, the prepotent response to muscle fatigue and being out of breath is to stop the physical exertion – in this experiment, participants had to exert control to keep going, and we hypothesised that this control would to some extent depend on the ability to use inner speech. Participants could use many different inner speech strategies and indeed claimed to do so in the self-talk questionnaire – regardless of which one, disrupting self-talk should disrupt control of the physical performance. The vast majority of our participants reported talking to themselves during exercise. We argue that under verbal interference, participants were less able to use their inner voice to focus their attention on the task demands and inhibit their propensity to slow down, and this had detrimental effects on their cycling performance. This is in line with previous dual-task literature suggesting that participants respond more impulsively (i.e., faster and with more errors) under verbal distraction conditions (Dunbar & Sussman, 1995; Nedergaard et al., 2022; Tullett & Inzlicht, 2010). In our experiment, the impulse would be to slow down.

One of the reasons why we decided to change the interference tasks from Experiment 1 to Experiment 2 was that the verbal interference task was substantially easier than the visuospatial interference task in Experiment 1. There was also no trade-off between cycling performance and interference task performance in Experiment 1, perhaps indicating that the interference tasks were not demanding enough. The issue with the difference in difficulty between the verbal and the visuospatial interference tasks was not quite solved in Experiment 2, although neither was at ceiling (in contrast to Experiment 1 where verbal interference task performance was near-perfect). To assess potentially problematic trade-off effects in more detail, we examined whether interference task condition and cycling performance predicted interference task performance. We found that the verbal interference task was indeed easier than the visuospatial interference task and that interference task performance decreased with increased cycling performance. However, there was no significant interaction between cycling performance and interference task condition, indicating that the trade-off was the same between interference task conditions. The absence of an interaction effect makes a direct comparison between their effects on cycling performance more reliable. The fact that the verbal interference task was easier than the visual indicates that we may still be underestimating the effect size of the direct comparison.

6.2. Effects of self-talk in sport

The present results provide an important additional perspective to the discussion on the effects of self-talk in sport. Existing dual-task studies investigating the involvement of cognitive functions in sport were ill-suited to answering our present questions as they were not designed to test verbal involvement specifically. Intervention studies on endurance sport have found that self-talk helps improve performance (Barwood et al., 2015; Blanchfield et al., 2014; Hamilton et al., 2007; Hatzigeorgiadis et al., 2018; McCormick et al., 2018; Schüler & Langens, 2007; P. J. Wallace et al., 2017). Because of the design of most of these intervention studies, it has not been possible to conclude that the self-talk interventions directly caused performance improvement – it could also simply be the case that undergoing any intervention helped, regardless of the content. The present study provides support for the claim that self-talk indeed has a direct causal role in performance.

A natural way to follow up on the present study would be to examine the role of inner speech in real endurance sports situations (such as marathons, triathlons, etc.) where it may be even more important how athletes talk to themselves. There is convincing evidence that marathon runners, for example, *believe* that self-talk helps them perform better (McCormick et al., 2018; Nedergaard et al., 2021; Schüler & Langens, 2007; Van Raalte et al., 2015) but evidence from interventions concerning whether it actually helps is mixed. As athletes generally differ in what kinds of self-talk helps them based on the type of sport (Theodorakis et al., 2000), their level of expertise (Nedergaard et al., 2021), and whether the setting is competition or training (Hatzigeorgiadis et al., 2014), we would expect interference to be differentially disruptive as well. For example, novices appear to benefit more from self-talk which yields the prediction that they would be more adversely affected by verbal interference (Nedergaard et al. 2021).

6.3. Limitations and future directions

The dual-task interference paradigm employed in the present study provides a promising avenue for future research in sport psychology. We were interested in spontaneous self-talk and thus did not ask our participants to say specific words or phrases to themselves the way it is usually done in intervention studies (Latinjak et al., 2019), but studies with a combination of self-talk training and verbal interference hold much potential. If one is interested in the effects of inner speech on behaviour and more particularly effects of the form and content of inner speech, it is informative to combine methods down-regulating language (such as verbal interference) with methods up-regulating language (such as self-talk training) (Nedergaard et al., 2022). Studies designed to inhibit linguistic processes, such as the present one, leave un-answered questions about what it is about inner speech that helps. Studies designed to increase specific ways of using language such as self-talk interventions are conversely limited in the causal claims they can make. The present study also contributes to the dual-task interference literature more generally by comparing effects of different types of interference (memory and continuous 2-back matching in this case). The fact that our continuous interference tasks yielded larger effects than the one-off memory interference tasks will be relevant for the choice of interference type in future studies.

Given that inner speech by definition is a hidden process which is primarily accessible through the verbal report of the speaker, it is potentially problematic that participants' self-report did not seem to be related to their physical performance or the degree to which they were affected by verbal interference. We expected that participants who said self-talk usually helps their performance would be more adversely affected by verbal interference than participants who said self-talk usually does not help their performance. There are four potential explanations for this mismatch: 1) Participants either do not remember accurately or are not aware of the actual relationship between the way they talk to themselves and the way they perform, 2) There are real relationships but we did not have a large or diverse enough sample to detect them, 3) The relationship differs depending on the type of exercise, where our study may not be prototypical, or 4) Participants are right about the usual effects of self-talk on their performance but the verbal interference in the present study did not in fact target self-talk. We believe that the first explanation is most likely given that introspective measures about causal explanations (such as 'talking to myself makes me perform better') are notoriously unreliable (Nisbett & Wilson, 1977). Reports are more likely to be accurate if participants are asked to simply report their experiences without attempting to theorise about the causes of their behaviour (Berger et al., 2016; Johansson et al., 2005; Petitmengin et al., 2013). This is in line with previous findings that questionnaire measures of inner speech have low correlation with more direct measures like experience sampling (Alderson-Day & Fernyhough, 2015a; Hurlburt et al., 2022).

7. CONCLUSION

The present study tested cycling performance during high-intensity 1-minute intervals under verbal and visuospatial interference conditions across two experiments using different kinds of interference tasks. While both interference conditions affected cycling performance negatively compared with a control condition, verbal interference was significantly worse than the control condition in Experiment 1 and worse than both the control and the visuospatial interference conditions in Experiment 2. Neither cycling performance nor degree of verbal interference seemed to have a consistent relationship with self-reported self-talk frequency or efficacy. Together, our experiments indicate that the inner voice plays an important role in the top-down control of physical performance.

Acknowledgments

This research was funded by an Interacting Minds Centre seed (2020) and the first author's PhD grant at the Faculty of Arts, Aarhus University. The authors thank Dan Mønster, Lasse Lui Frandsen, Solveig Topp, and Kim Topp for assistance with the experimental setup and materials for both experiments.

8. REFERENCES

Alderson-Day, B., & Fernyhough, C. (2015a). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00517

Alderson-Day, B., & Fernyhough, C. (2015b). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

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), 641–657. https://doi.org/10.1037/0096-3445.130.4.641

Baldo, J., Dronkers, N., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language*, *92*(3), 240–250. https://doi.org/10.1016/j.bandl.2004.06.103

Barwood, M. J., Corbett, J., Wagstaff, C. R., McVeigh, D., & Thelwell, R. C. (2015). Improvement of 10-km time-trial cycling with motivational self-talk compared with neutral self-talk. *International Journal of Sports Physiology and Performance*, *10*(2), 166–171.

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

Bek, J., Blades, M., Siegal, M., & Varley, R. (2009). Linguistic Processes in Visuospatial Representation: Clarifying Verbal Interference Effects. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*.

Bek, J., Blades, M., Siegal, M., & Varley, R. (2013). Dual-Task Interference in Spatial Reorientation: Linguistic and Nonlinguistic Factors. *Spatial Cognition & Computation*, *13*(1), 26–49. https://doi.org/10.1080/13875868.2011.590622

Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D. C., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. *PloS One*, *13*(2), e0192691.

Berger, C. C., Dennehy, T. C., Bargh, J. A., & Morsella, E. (2016). Nisbett and Wilson (1977) revisited: The little that we can know and can tell. *Social Cognition*, *34*(3), 167.

Biese, K. M., Pietrosimone, L. E., Andrejchak, M., Lynall, R. C., Wikstrom, E. A., & Padua, D. A. (2019). Preliminary Investigation on the Effect of Cognition on Jump-Landing Performance Using a Clinically Relevant Setup. *Measurement in Physical Education and Exercise Science*, *23*(1), 78–88. https://doi.org/10.1080/1091367X.2018.1518875

Blakely, M. J., Kemp, S., & Helton, W. S. (2016). Volitional Running and Tone Counting: The Impact of Cognitive Load on Running Over Natural Terrain. *IIE Transactions on Occupational Ergonomics and Human Factors*, *4*(2–3), 104–114. https://doi.org/10.1080/21577323.2015.1055864

Blanchfield, A. W., Hardy, J., De Morree, H. M., Staiano, W., & Marcora, S. M. (2014). Talking yourself out of exhaustion: The effects of self-talk on endurance performance. *Med Sci Sports Exerc*, *46*(5), 998–1007.

Borchers, H. W. (2021). *pracma: Practical Numerical Math Functions*. https://CRAN.R-project.org/package=pracma

Brick, N. E., MacIntyre, T. E., & Campbell, M. J. (2016). Thinking and Action: A Cognitive Perspective on Self-Regulation during Endurance Performance. *Frontiers in Physiology*, *7*. https://doi.org/10.3389/fphys.2016.00159

Bryck, R. L., & Mayr, U. (2005). On the role of verbalization during task set selection: Switching or serial order control? *Memory & Cognition*, *33*(4), 611–623. https://doi.org/10.3758/BF03195328

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition,* *2*(1): 16, pp. 1–38. DOI: https://doi.org/10.5334/joc.72

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

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

Craig, N. P., Pyke, F. S., & Norton, K. I. (1989). Specificity of test duration when assessing the anaerobic lactacid capacity of high-performance track cyclists. *International journal of sports medicine, 10*(04), 237-242.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dunbar, K., & Sussman, D. (1995). Toward a Cognitive Account of Frontal Lobe Function: Simulating Frontal Lobe Deficits in Normal Subjects. *Annals of the New York Academy of Sciences*, *769*(1 Structure and), 289–304. https://doi.org/10.1111/j.1749-6632.1995.tb38146.x

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. https://doi.org/10.1016/S0749-596X(02)00511-9

Epling, S. L., Blakely, M. J., Edgar, G. K., Russell, P. N., & Helton, W. S. (2018). Memory impairment during a climbing traverse: Implications for search and rescue climbing. *Experimental Brain Research*, *236*(11), 3043–3052. https://doi.org/10.1007/s00221-018-5362-8

Green, A. L., & Helton, W. S. (2011). Dual-task performance during a climbing traverse. *Experimental Brain Research*, *215*(3–4), 307–313. https://doi.org/10.1007/s00221-011-2898-2

Hamilton, R. A., Scott, D., & MacDougall, M. P. (2007). Assessing the effectiveness of self-talk interventions on endurance performance. *Journal of Applied Sport Psychology*, *19*(2), 226–239.

Hatzigeorgiadis, A., Bartura, K., Argiropoulos, C., Comoutos, N., Galanis, E., & D. Flouris, A. (2018). Beat the heat: Effects of a motivational self-talk intervention on endurance performance. *Journal of Applied Sport Psychology*, *30*(4), 388–401.

Hatzigeorgiadis, A., Galanis, E., Zourbanos, N., & Theodorakis, Y. (2014). Self-talk and Competitive Sport Performance. *Journal of Applied Sport Psychology*, *26*(1), 82–95. https://doi.org/10.1080/10413200.2013.790095

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., & Theodorakis, Y. (2011). Self-Talk and Sports Performance: A Meta-Analysis. *Perspectives on Psychological Science*, *6*(4), 348–356. https://doi.org/10.1177/1745691611413136

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective Interference with Verbal Short-Term Memory for Serial Order Information: A New Paradigm and Tests of a Timing-Signal Hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1307–1334. https://doi.org/10.1080/02724980244000747

Hurlburt, R. T., Heavey, C. L., Lapping-Carr, L., Krumm, A. E., Moynihan, S. A., Kaneshiro, C., Brouwers, V. P., Turner, D. K., & Kelsey, J. M. (2022). Measuring the Frequency of Inner-Experience Characteristics. *Perspectives on Psychological Science*, *17*(2), 559–571. https://doi.org/10.1177/1745691621990379

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science*, *310*(5745), 116–119. https://doi.org/10.1126/science.1111709

Kirschenbaum, D. S. (1987). Self-regulation of sport performance: *Medicine & Science in Sports & Exercise*, *19*(Supplement), S114-S120. https://doi.org/10.1249/00005768-198710001-00003

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

Latinjak, A. T., Hatzigeorgiadis, A., Comoutos, N., & Hardy, J. (2019). Speaking clearly . . . 10 years on: The case for an integrative perspective of self-talk in sport. *Sport, Exercise, and Performance Psychology*, *8*(4), 353–367. https://doi.org/10.1037/spy0000160

Makowski, D. (2018). The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Martin, J. C., Davidson, C. J., & Pardyjak, E. R. (2007). Understanding sprint-cycling performance:

the integration of muscle power, resistance, and modeling. *International journal of sports physiology and performance, 2*(1), 5-21.

McCormick, A., Meijen, C., Anstiss, P. A., & Jones, H. S. (2019). Self-regulation in endurance sports: Theory, research, and practice. *International Review of Sport and Exercise Psychology*, *12*(1), 235–264. https://doi.org/10.1080/1750984X.2018.1469161

McCormick, A., Meijen, C., & Marcora, S. (2018). Effects of a motivational self-talk intervention for endurance athletes completing an ultramarathon. *The Sport Psychologist*, *32*(1), 42–50.

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. https://doi.org/10.1016/j.concog.2021.103102

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*. https://doi.org/10.3758/s13423-022-02144-7

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Petitmengin, C., Remillieux, A., Cahour, B., & Carter-Thomas, S. (2013). A gap in Nisbett and Wilson's findings? A first-person access to our cognitive processes. *Consciousness and Cognition*, *22*(2), 654–669. https://doi.org/10.1016/j.concog.2013.02.004

Piccardi, L., Bocchi, A., Palmiero, M., Boccia, M., D'Amico, S., & Nori, R. (2020). Chatting While Walking Does Not Interfere with Topographical Working Memory. *Brain Sciences*, *10*(11), 811. https://doi.org/10.3390/brainsci10110811

Quer, G., Gouda, P., Galarnyk, M., Topol, E. J., & Steinhubl, S. R. (2020). Inter-and intraindividual

variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: Retrospective, longitudinal cohort study of 92,457 adults. *Plos one, 15*(2), e0227709.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reimers, A. K., Knapp, G., & Reimers, C.-D. (2018). Effects of exercise on the resting heart rate: A systematic review and meta-analysis of interventional studies. *Journal of Clinical Medicine*, *7*(12), 503.

Ren, X., Wang, T., & Jarrold, C. (2016). Individual Differences in Frequency of Inner Speech: Differential Relations with Cognitive and Non-cognitive Factors. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01675

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*(6), 977–986. https://doi.org/10.3758/BF03209345

Schüler, J., & Langens, T. A. (2007). Psychological Crisis in a Marathon and the Buffering Effects of Self-Verbalizations. *Journal of Applied Social Psychology*, *37*(10), 2319–2344. https://doi.org/10.1111/j.1559-1816.2007.00260.x

Schweizer, G., & Furley, P. (2016). Reproducible research in sport and exercise psychology: The role

of sample sizes. *Psychology of Sport and Exercise, 23*, 114-122.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stets, A., Smith, S. L., & Helton, W. S. (2020). Dual-Task Interference Between Swimming and Verbal Memory. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(7), 1132–1140. https://doi.org/10.1177/0018720819871743

Świątkowski, W., & Dompnier, B. (2017). Replicability Crisis in Social Psychology: Looking at the Past to Find New Pathways for the Future. *International Review of Social Psychology*, *30*(1), 111. https://doi.org/10.5334/irsp.66

Talarico, M. K., Lynall, R. C., Mauntel, T. C., Weinhold, P. S., Padua, D. A., & Mihalik, J. P. (2017). Static and dynamic single leg postural control performance during dual-task paradigms. *Journal of Sports Sciences*, *35*(11), 1118–1124. https://doi.org/10.1080/02640414.2016.1211307

Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, *37*(1), 153–156. https://doi.org/10.1016/S0735-1097(00)01054-8

Theodorakis, Y., Weinberg, R., Natsis, P., Douma, I., & Kazakas, P. (2000). The effects of motivational versus instructional self-talk on improving motor performance. *The Sport Psychologist*, *14*(3), 253–271.

Tod, D., Hardy, J., & Oliver, E. (2011). Effects of Self-Talk: A Systematic Review. *Journal of Sport and Exercise Psychology*, *33*(5), 666–687. https://doi.org/10.1123/jsep.33.5.666

Torchiano, M. (2020). *effsize: Efficient Effect Size Computation*. https://doi.org/10.5281/zenodo.1480624

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, *135*(2), 252–256. https://doi.org/10.1016/j.actpsy.2010.07.008

Van Boxtel, G. J. M., & et al. (2021). *gsignal: Signal processing*. https://github.com/gjmvanboxtel/gsignal

Van Raalte, J. L., Morrey, R. B., Cornelius, A. E., & Brewer, B. W. (2015). Self-talk of marathon runners. *The Sport Psychologist*, *29*(3), 258–260.

Van Raalte, J. L., Vincent, A., & Brewer, B. W. (2016). Self-talk: Review and sport-specific model. *Psychology of Sport and Exercise*, *22*, 139–148. https://doi.org/10.1016/j.psychsport.2015.08.004

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. https://doi.org/10.1044/2017_JSLHR-S-16-0376

Wallace, P. J., McKinlay, B. J., Coletta, N. A., Vlaar, J. I., Taber, M. J., Wilson, P. M., & Cheung, S. S. (2017). Effects of motivational self-talk on endurance and cognitive performance in the heat. *Medicine & Science in Sports & Exercise*, *49*(1), 191–199.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Woodham, A., Billinghurst, M., & Helton, W. S. (2016). Climbing With a Head-Mounted Display: Dual-Task Costs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 452–461. https://doi.org/10.1177/0018720815623431

Zourbanos, N., Hatzigeorgiadis, A., Chroni, S., Theodorakis, Y., & Papaioannou, A. (2009). Automatic self-talk questionnaire for sports (ASTQS): Development and preliminary validation of a measure identifying the structure of athletes' self-talk. *The Sport Psychologist*, *23*(2), 233–251.

Article III

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

https://doi.org/10.1037/xhp0001112

"Stay Focused!": The Role of Inner Speech in Maintaining Attention During a Boring Task

Johanne Nedergaard, Joshua Charles Skewes, and Mikkel Wallentin
Department of Linguistics, Cognitive Science and Semiotics, Aarhus University

Is inner speech involved in sustaining attention, and is this reflected in response times for stimulus detection? In Experiment 1, we measured response times to an infrequently occurring stimulus (a black dot appearing at 1–3 min intervals) and subsequently asked participants to report on the character of their inner experience at the time the stimulus appeared. Our main preregistered hypothesis was that there would be an interaction between inner speech and task relevance of thought with reaction times being the fastest on prompts preceded by task-relevant inner speech. This would indicate that participants could use their inner voice to maintain performance on the task. With generalized linear mixed-effects models fitted to a gamma distribution, we found significant effects of task relevance but no interaction with inner speech. However, using a hierarchical Bayesian analysis method, we found that trials preceded by task-relevant inner speech additionally displayed lower standard deviation and lower mode (independently of the main effect of task relevance), suggestive of increased processing efficiency. Due to deviations from the preregistered sampling and analysis procedures, we replicated our findings in Experiment 2. Our results add support to the hypothesis that inner speech serves a functional role in top-down attentional control.

Public Significance Statement
This study suggests that reaction time performance on a boring task demanding nothing but sustained attention benefits from task-relevant inner experience generally and task-relevant *inner speech* specifically. This indicates that inner speech is employed as a tool for behavioral control in this domain.

Keywords: sustained attention, behavioral control, inner speech, mind-wandering

Supplemental materials: https://doi.org/10.1037/xhp0001112.supp

Theories of inner speech have proposed several different cognitive functions, among which are as a mnemonic device (Emerson & Miyake, 2003), for speech processing (Jacquemot & Scott, 2006), and for behavioral and attentional control (Alderson-Day & Fernyhough, 2015b; Morin et al., 2011). In this study, we investigate the behavioral control function of inner speech. Experience sampling studies and questionnaire studies have suggested that people often talk to themselves in a self-regulatory way, although these studies provide little evidence as to whether self-regulatory inner speech actually has an effect on behavior (Alderson-Day et al., 2018; Alderson-Day & Fernyhough, 2015a; Morin et al., 2011, 2018; Uttl et al., 2011). Support for this assumption has come from sport psychology research (Hatzigeorgiadis et al., 2011; Nedergaard et al., 2021; Tod et al., 2011) where participants are often trained to talk to themselves in a specific way, with behavioral outcomes being relatively simple to measure (usually enhanced endurance performance or motor control). These studies, however, have methodological challenges such as small convenience samples and lack of active control groups. In the present study, we wanted to examine the effects of naturally occurring inner speech that was either task-relevant or task-irrelevant in a task that was designed to be uneventful and tedious. Thus, we tested the role of inner speech in the kind of self-control involved in sustained attention.

Inner Speech and Behavioral Control

Self-talk appears to play an important role in the acquisition, maintenance, and execution of physical skills (Hatzigeorgiadis et al., 2011; Tod et al., 2011). It seems to be the case that task-relevant, focused self-talk is recruited under circumstances that are highly demanding, either because the athlete is learning a new sport (Zourbanos et al., 2013), competing against others (Dickens et al., 2018; Thibodeaux & Winsler, 2018; Van Raalte et al., 2000) or under high intensity (Aitchison et al., 2013; Nedergaard et al., 2021). In the first study to test self-talk and physical performance with a dual-task interference paradigm (Nedergaard, Wallentin, & Lupyan, 2022), Nedergaard, Christensen, and Wallentin (2022) found that participants who cycled on an exercise bike while engaged in simultaneous verbal interference were slower than when they were cycling without interference. In dual-task paradigms such as this, a specific negative effect of verbal interference is taken to mean that participants under normal circumstances benefit (in this case in the form

Johanne Nedergaard https://orcid.org/0000-0001-6381-2901
All experiment code and data can be accessed online at the Open Science Forum via this link: https://osf.io/jgx7m/
Correspondence concerning this article should be addressed to Johanne Nedergaard, Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark. Email: nedergaard@cc.au.dk

of better cycling performance) from being able to talk to themselves (see Nedergaard, Christensen, & Wallentin, 2022, for a comprehensive review of the verbal interference literature).

Aside from motor control, it also appears that people recruit internal language for impulse control more generally—to stay focused on a task that is tedious or to refrain from making inappropriate responses. For example, Tullett and Inzlicht (2010) tested inhibitory control in a Go/No-Go task in combination with a verbal interference paradigm and found that when participants were engaged in articulatory suppression, they were more prone to impulsive responding as indicated by a greater tendency to make a "Go" response. The authors interpreted their findings to mean that people usually use their inner voice to inhibit impulsive responding. The evidence from cognitive psychology parallels that from sport psychology in that inner speech appears to be especially recruited under challenging circumstances, when learning new skills or when a high degree of top-down control is necessary (Emerson & Miyake, 2003; Kray et al., 2008).

In situations demanding top-down control of attention, mind-wandering is associated with failures to monitor task performance, thus leading to more errors (Smallwood et al., 2007). The literature on mind-wandering has generally not been concerned with the specific modality in which inner experience takes place, but rather whether it is task-relevant or not. "Inner experience" in this context refers to subjectively experienced mental phenomena such as feelings, desires, thoughts, reasonings, and decisions that are accessible to verbal report. Interestingly, response times in a sustained attention task (the Metronome Response Task) showed more variability prior to self-reported mind-wandering compared to response times prior to self-reported task-relevant inner experience (Seli, Carriere, et al., 2013; Seli, Cheyne, & Smilek, 2013). As in our present study, Seli and colleagues recruited participants online. In the present study, we attempt to explicate the mechanisms underlying the increased variability in reaction time associated with task-irrelevant inner experience. We do this by putting more emphasis on the format of inner experience. Previous mind-wandering studies have also found that performance is connected to task-relevant or task-irrelevant thought. For example, Welhaf et al. (2020) also probed participants' thoughts after different response time tasks (e.g. Stroop and Flanker tasks) and found that "task-unrelated thought" (reports of thinking about "everyday things," "current state of being," "personal worries," "daydreams," "external environment," or "other" thoughts) correlated more strongly with participants' worst response times than with best or mean response times. No distinction between verbal and non-verbal thoughts was made.

Measuring Inner Experience

It appears that inner speech and task-relevant experience are related to better motor and attentional control. But how can we know what the content of experience is? In recent years, one method, in particular, has received considerable attention: Descriptive Experience Sampling. This method has participants carry a beeper and note down the format and content of their internal experience at random points during the day. Using this method and others like it, we generally see five main types of internal experience: inner voice, inner seeing, sensory awareness, unsymbolized thinking, and feelings. These each appear to occur in approximately 25% percent of sampled experiences, and multiple experience types can occur at the same time (Heavey & Hurlburt, 2008). It is worth noting, however, that the reporting of the phenomena of inner experience is highly susceptible to the way the questions are phrased. In a different experience sampling study, Uttl et al. (2012) for example found that inner speech occurred in 60% of sampled moments, potentially because they did not allow for other types of inner experience. Interestingly, questionnaire-based methods appear to overestimate the frequency of all experience types compared with Descriptive Experience Sampling (Hurlburt et al., 2022). Overall, inner speech appears to be self-centered (Morin et al., 2011) and to serve problem-solving, planning, motivational, mnemonic, and evaluation functions (Morin et al., 2018). Nevertheless, given the controversy surrounding the reliability and validity of introspection in this area, the relationship between inner speech and behavioral control still proves elusive.

The Present Study

In the present study, we investigate how people manage to stay focused on a task that does not demand anything but their visual attention. We were interested in whether the format and content of inner experience immediately before a reaction time prompt had any influence on the speed with which participants were able to respond to the prompt. In order to allow participants' minds to wander, we inserted relatively long breaks between stimuli. The experiment was intentionally boring to ensure that participants needed to exercise self-control to stay focused. This was particularly important as the experiment took place online, meaning that any drive to comply with the experimenter was greatly diminished. The design of the present experiment illustrates a novel way of measuring the relationship between inner experience and behavior. It avoids the resource-intensiveness of Descriptive Experience Sampling and alleviates the lack of reliability associated with questionnaire measures by being concurrent and non-theorizing (Ericsson & Simon, 1980).

After each trial, participants answered questions about their inner experiences. These questions (see Table 1) were inspired by Descriptive Experience Sampling research. We thus asked if their inner experience took the form of inner voice, inner seeing, unsymbolized thinking, feelings, or sensory awareness. Of these five, "unsymbolized thinking" is perhaps the most opaque—it is described by Heavey and Hurlburt (2008) as "Thinking a particular, definite thought without the awareness of that thought's being conveyed in words, images, or any other symbols" (p. 802). Inspired by the mind-wandering literature (Gonçalves et al., 2017; Maillet & Rajah, 2013; Mrazek et al., 2012; Smallwood et al., 2007), we also asked participants whether their experience was task-relevant or not, whether their experience was about past, present, or future, and whether they were aware of their experience before the prompt appeared. Previous studies (Smallwood & Schooler, 2015) occasionally differentiate between perceptually guided, on-task focus (e.g. focus on task requirements) and self-generated, task-related thoughts (thoughts about the task that are not about a focus on completing it, e.g. "this task is so boring!"). The former would plausibly appear in our study as task-relevant sensory awareness (i.e. focus on the visual stimulus of the fixation cross). The latter could be task-relevant experiences of any format.

If participants responded that their inner experience had a verbal quality, we asked them some additional questions inspired by Alderson-Day and colleagues and their research on inner speech and self-regulation (Alderson-Day et al., 2018). These additional questions were about how dialogic, condensed (i.e. experienced as

Table 1
The Questions Posed to Participants After Each Circle Prompt

Question (inner experience)	Options
Were your thoughts about the current task or not?	“Yes,” “No,” “I don’t know”
Were your thoughts about past, present or future?	“Past,” “Present,” “Future,” “I don’t know”
Were you aware of your own thoughts before you saw the circle?	“Yes,” “No,” “I don’t know”
How would you characterize your inner experience just before you saw the circle?	“Inner voice,” “Inner seeing,” “Unsymbolized thinking,” “Sensory awareness,” “Feelings”
Question (inner speech-specific)	Options
I was having a back and forth conversation in my head.	“Disagree,” “Partially disagree,” “Neither agree nor disagree,” “Partially agree,” “Agree”
My thinking was shortened compared to my normal, out-loud speech.	*Same as the above.*
I was having the experience of other people’s voices.	*Same as the above.*
I was evaluating my behavior using my inner speech.	*Same as the above.*

abbreviated or with missing syntactic or morphological elements compared to normal, out-loud speech—but with meaning retained), and evaluative their inner speech was, and whether they had the experience of other people’s voices.

We decided to collect data online for several reasons. First, it made it possible to recruit a larger sample than would have been possible for practical reasons in the laboratory. Second, participants recruited online are likely to represent a more diverse and representative group than participants at a behavioral laboratory at a university. Laboratory participants are generally highly skewed toward high socioeconomic status and high levels of education (Hartshorne, 2020), and studies conducted online are found to yield at least as good data as studies conducted in the laboratory (Hartshorne et al., 2019).

Our preregistered hypotheses were as follows (https://osf.io/stfn5):

H1: Task-relevant inner experience will generally be associated with faster reaction times to the prompt.

H2: Specifically, task-relevant inner speech will be associated with faster reaction times than other types of inner experience.

H3: The proportions of types of inner experience will resemble those found in other experience sampling studies.

H4 (exploratory): Self-regulatory inner speech may be more important as the experiment progresses. If this is the case, we predict an interaction between the inner speech factor and time with the difference between task-relevant inner speech and task-relevant other experience becoming more pronounced over time.

H5 (exploratory—see Registration 2): Response time variance will be lower for task-relevant inner speech trials.

Method: Experiment 1

Transparency and Openness Statement

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. All data, analysis code, and research materials are available at https://osf.io/jgx7m/. Data were analyzed using R, version 4.0.0 (R Core Team, 2022), and the package ggplot, version 3.2.1 (Wickham, 2016). This study’s design and its analysis were preregistered (https://osf.io/stfn5).

Participants

Power analysis conducted using the R package “simr” (Green & MacLeod, 2016) based on pilot data from 10 participants suggested that we needed to recruit 120 participants to be able to detect a 40 ms difference between task-relevant and task-irrelevant trials with a power of 94.00% (95% CI [83.45, 98.75]). The linear mixed model used in this power analysis was a model of task relevance predicting reaction time (as normally distributed) with random intercepts modeled for each participant. We did not initially test for our power to detect an interaction effect which a reviewer pointed out. We conducted this analysis post hoc and found that our power to detect an interaction effect of 40 ms between inner speech and task relevance was 44% [29.99, 58.75]. The model used for this power analysis was identical except that both task relevance, inner speech, and the interaction between them were included as predictors. The post hoc power analysis suggested that we would need 220 participants to detect an interaction effect of 40 ms with a power of 79% [69.71, 86.51]. This power analysis provides further justification for re-analyzing the data using Bayesian methods (see Unregistered Analyses section) because such methods let us focus on how confident we can be about the size of the effects rather than their likelihood of being detected by significance tests. We will discuss the Bayesian analysis in more detail after first reporting the preregistered analyses. We recruited all participants through the online platform Prolific and required that they had English as their first language and access to a desktop browser. These factors necessarily constrain the generalizability of our results. Native English speakers with access to a desktop browser may not be universally representative as a sample. Nevertheless, the language constraint was necessary to ensure that participants understood the instructions, and the equipment constraint was necessary to minimize technical errors during the experiment. Recruiting participants on an online platform such as Prolific yields a wider demographic range than recruiting from a university setting so we believe our results are relatively generalizable.

In the first round of data collection, not enough participants reported all four combinations of task-relevant experience and inner speech (task-relevant inner speech, task-irrelevant inner speech, task-relevant non-inner speech, and task-irrelevant non-inner speech). In fact, only 77 out of 120 reported all four kinds of experience. This meant that we did not reach the threshold established by our power analysis, and we thus needed to collect data from more participants. We submitted another preregistration before the second round of data collection (https://osf.io/jb3c8). Below we report results from the first and second rounds combined for brevity and clarity. See the online supplemental materials for analyses separated into the first and second rounds. See Table 2 for the demographic data for participants in both rounds of data collection. Ethical approval for the experiment was obtained through the Institutional Review Board at Aarhus University.

Materials

The experiment was custom-written in JavaScript using the jsPsych library (De Leeuw, 2015). In order to ensure that participants stayed focused on the task, we recorded their browser interactions (a feature built into jsPsych)—when they left and entered full-screen mode, when they left and entered the tab the experiment was in ("blur" and "focus," respectively), the index of the trial they were in when the browser event occurred, and the time since the experiment started. The black dot prompt that participants had to respond to had a diameter of approximately 20% of the screen width (gray background).

Procedure

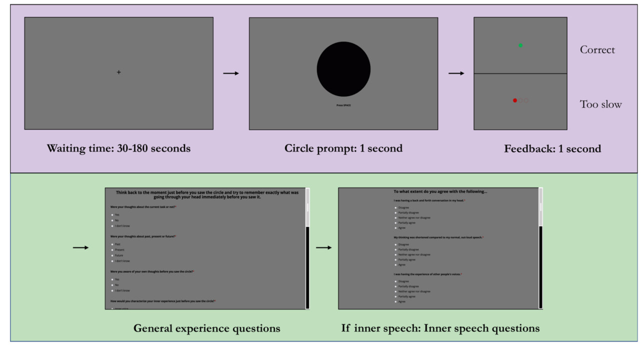
Participants first saw an instruction screen that informed them what the experiment was about, encouraged them to read an attached informed consent sheet, and instructed them that they had to keep their gaze fixed on a fixation cross during waiting periods and that they would only have one second to respond to each prompt. Participants then went through a short set of three practice trials with shorter interstimulus intervals than the real experiment (a few seconds instead of a few minutes) to get used to responding to the circle prompt. In the real experiment, each participant responded to eight circle prompts. The order of the interstimulus intervals was randomized for each participant (30, 50, 60, 70, 120, 120, 150, 180 s). If they failed to respond to a circle prompt within 1 s, they saw a feedback screen showing the number of missed prompts they had accumulated (in the form of red dots). See Figure 1 for a schematic of the experiment progression. After the circle prompt and the feedback, participants were asked about their inner experience (see Table 1 for the full set of questions). The specific prompt was "Think back to the moment just before you saw the circle and try to remember exactly what was going through your head immediately before you saw it." The data were collected in June, July, and September 2021.

When the participant had responded to all eight circle prompts, they were also asked whether they talked to themselves to stay focused throughout the experiment.

Table 2
Demographic Data for All Participants in Both Rounds of Data Collection

N (after exclusions)	212
Excluded (failed to respond to three prompts)	22
Female	117
Male	92
Data from Prolific expired	3
Median age (range)	31 years (18–83)
Median time spent (range)	19 min 5 s (15 min 22 s–62 min 25 s)

Figure 1
Illustration of the Experiment Progression



Note. See the online article for the color version of this figure.

Results: Experiment 1

Because the study was conducted online, we expected a high degree of variability in responses as we could not control participants' immediate environment, and they were likely to be a more heterogenous sample than the usual university students. We report first the preregistered analyses and then unregistered analyses. In all modeling, we excluded trials where participants answered "I don't know" to the task relevance questions as some cells would otherwise have less than five instances.

Descriptive Statistics

We report both the results from the first and second data collection rounds combined (see above for details). In the interest of transparency, separate statistics are reported in the online supplemental materials.

Reaction Time

As detailed in the preregistration, we excluded trials with reaction times below 200 ms and trials from participants who missed three prompts (22 participants). Aside from these participants whose data were not included, we excluded 75 trials where participants were too slow to respond (RT < 1,000 ms, 4.6% of all trials). Consistent with our conception of reaction times as waiting times, the reaction time data were gamma rather than normally distributed (see Figure 2).

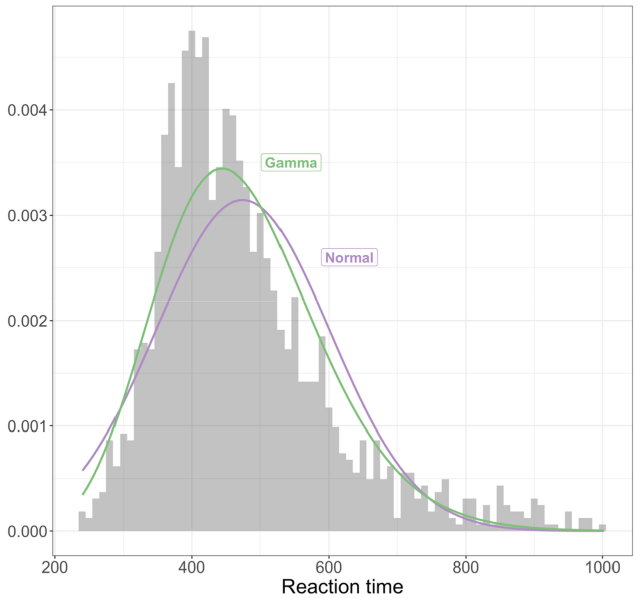
Participants had a mean reaction time of 474.15 ms ($SD =$ 126.83) and a median reaction time of 448.23 ms, supporting the assumption that reaction times were positively skewed and thus followed a gamma rather than a normal distribution.

The fitted models' log-likelihoods are as follows: Normal = −10,149.80; gamma = −10,013.54. The higher the log-likelihood, the better the fit, confirming that the gamma distribution fits the data best. For that reason, as well as for theoretical reasons (Lo & Andrews, 2015), we use a gamma distribution for the remaining analyses.

Experience Questions

See Table 3 and Figure 3 for proportions of reported inner experience types. Chi-square tests suggested that task relevance and

Figure 2
Fitted Gamma and Normal Distributions for Reaction Times Across All Participants



Note. The gamma distribution models a continuous distribution with two parameters (shape and rate) which is often used to model wait times and other phenomena that are always positive and skewed. When the shape parameter is >1, the distribution is positively skewed. The normal distribution is symmetric and models a continuous distribution with two parameters (mean and standard deviation). See the online article for the color version of this figure.

experience type were not independent when excluding trials where participants responded “I don’t know” to the task relevance question ($\chi^2 = 76.72$, $df = 4$, $p < .001$).

Inner Speech Questions

Participants reported that their inner voice was condensed, evaluative and dialogic (≍50% of the trials), but rarely the voice of somebody else (≍10% of the trials) (Figure 4, left). This pattern of responses is comparable to a previous study (Figure 4, right) using similar items (Alderson-Day et al., 2018), despite the differences in Likert scales (five-point in ours and seven-point in the original VISQ-R study). In our sample, 24.1% of participants never reported experiencing inner speech.

When asked at the end of the experiment whether they had talked to themselves to stay focused during the experiment, 164 participants (77.4%) said that they had and 48 participants (22.6%) said that they had not.

Task Relevance as a Predictor of Reaction Time

We fitted a gamma generalized linear mixed model with a log link function with reaction time predicted by task relevance. The model included random intercepts for each participant. Task relevance significantly predicted reaction time ($\beta = -0.02$, $SE = 0.01$, $p = .037$) with trials preceded by task-relevant experience having a faster reaction time than trials preceded by task-irrelevant experience. As the coefficients are in log space, we back-transformed them for interpretability and found that reported task-relevant inner experience was associated with a 2% decrease in reaction time. To check that the effect was not just driven by a few individuals, we examined how many of the participants were faster with task-relevant experience. This was the case for 64.8% of the participants who reported both kinds of trials ($N = 145$).

Task Relevance and Inner Speech as Combined Predictors of Reaction Time

We fitted a gamma generalized linear mixed model with a log link function with reaction time predicted by task relevance, inner speech, and the interaction between them. The model included random intercepts for each participant. None of the predictors were statistically significant in this model (all $p > .264$). It is important to take particular note of the fact that task relevance was no longer a significant predictor when combined with inner speech in this model. This suggests that the effect of task relevance was at least partially explained by its interaction with inner speech (see Bayesian models below).

The Effect of Trial Progression

We fitted a gamma generalized linear mixed model with a log link function with reaction time predicted by task relevance and trial progression. The model failed to converge with random intercepts for each participant so we did not include them. Trial significantly predicted reaction time ($\beta = 0.01$, $SE = 0.005$, $p = .039$) with later trials being associated with slower reaction times. As the coefficients are in log space, we back-transformed them for interpretability and found that each increase in trial was associated with a 1.1% increase in reaction time. There was no interaction between trial progression and task relevance ($p = .276$).

Unregistered Analyses

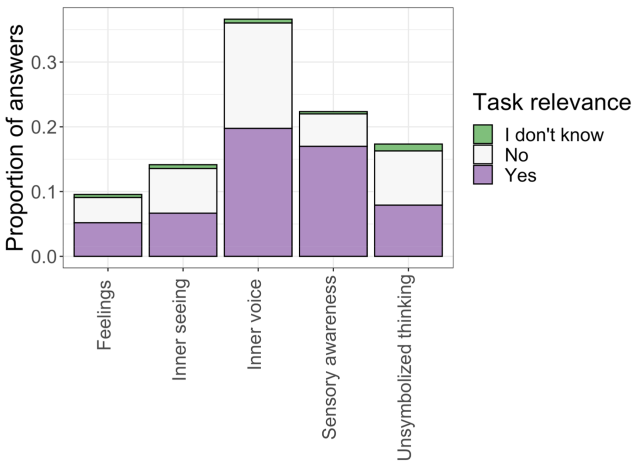
During analyses, we noticed an interesting pattern in the distributions of the results which seemed to be related to the spread of the

Table 3
Reported Types of Inner Experience in Percentages Across All Prompts (Eight Per Participant)

Experience type	Task-relevant (count)	Task-irrelevant (count)	“I don’t know” responses (count)	Percentage of total samples
Feelings	88	66	8	9.44%
Inner seeing	113	117	10	13.99%
Inner voice	335	276	10	36.19%
Sensory awareness	288	85	6	22.09%
Unsymbolized thinking	134	142	18	17.13%

Figure 3

Visualization of Reported Types of Inner Experience and Whether They Were Relevant to the Task or Not Across All Prompts (Eight Per Participant)



Note. The difference in proportions between task-relevant/task-irrelevant thoughts was significant. The main difference seems to be sensory awareness, which is reported more frequently for task-relevant thought. See the online article for the color version of this figure.

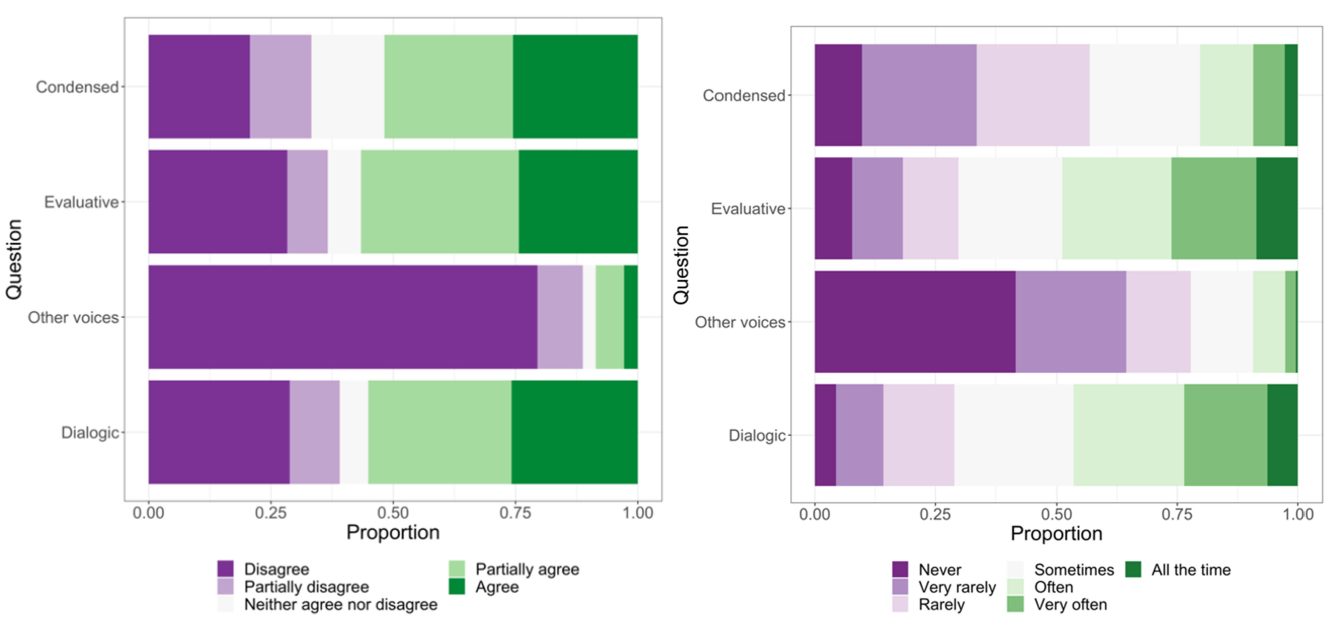
data rather than necessarily the central tendencies (see Figure 5). It appeared that task-relevant inner speech trials were different from the types of trials in a way that was not captured by our preregistered analyses. Specifically, it appeared that not only might the peak of the reaction time distribution be shifted as an effect of task relevance, but also that the spread or variability in reaction times might be different, and that task-relevant inner speech might also increase the precision of response times (Figure 5). For this reason, we constructed a new analysis that simultaneously modeled both changes in mode and changes in precision of the reaction time distribution. To do this we used hierarchical Bayesian modeling. This has the added advantage of avoiding some of the convergence problems in the preregistered analysis. In addition, the power analysis we conducted suggested that we did not have sufficient power to detect an interaction effect between inner speech and task relevance. The Bayesian analysis ameliorates this problem, by allowing us to report an inference that is not dependent on assumptions about the long-run likelihood that a true effect can be detected by a significance test.

Hierarchical Bayesian Models

Conducting hierarchical Bayesian analyses allowed us to model each individual participant's reaction times as gamma distributions and let us test for differences in both variation and central tendency instead of just central tendency. We compiled the models detailed below using JAGS (Just Another Gibbs Sampler) which uses MCMC (Markov chain Monte Carlo) sampling (Plummer, 2003). Our JAGS models were implemented in R using the R2JAGS package (Su & Masanao, 2021). We were interested in the differences between modes and standard deviations for trials following task-relevant versus task-irrelevant experience (main effect of task relevance) and whether task-relevant inner speech reduced both the mode and the standard deviation of the reaction time distribution. For all models, the full model specifications can be found in the online supplemental materials. The different effects of interest

Figure 4

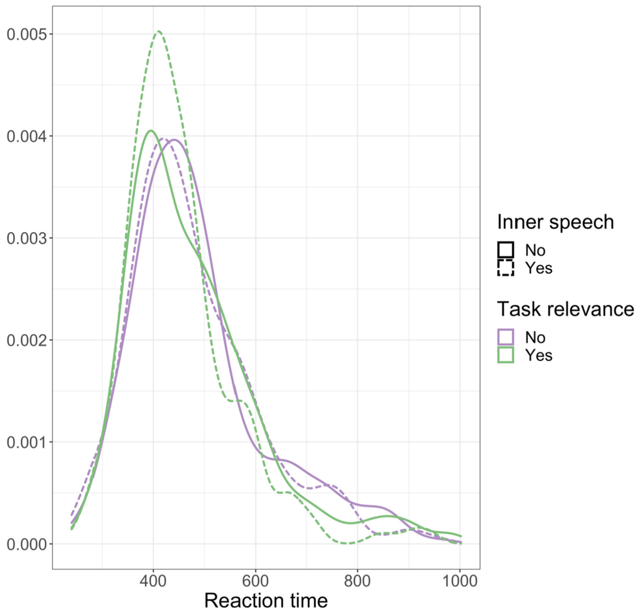
On the Left are the Answers to the Inner Speech Questions in the Present Experiment



Note. On the right are the answers extracted from VISQ-R (Alderson-Day et al., 2018). Note that we had a five-point Likert scale while Alderson-Day et al. had a seven-point Likert scale. See the online article for the color version of this figure.

Figure 5

Visualization of the Densities of Reaction Times in the Four Combinations of Task Relevance and Inner Speech



Note. Note that the mode of task-relevant trials is lower than those of task-irrelevant trials. Also, note that the width of the task-relevant inner speech distribution seems to be narrower than those of the other conditions. This suggests that trials preceded by task-relevant inner speech is associated with greater RT precision. We used hierarchical Bayesian modeling to investigate this. See the online article for the color version of this figure.

(main effect of task relevance and the interaction between task relevance and inner speech) were tested by defining different contrasts.

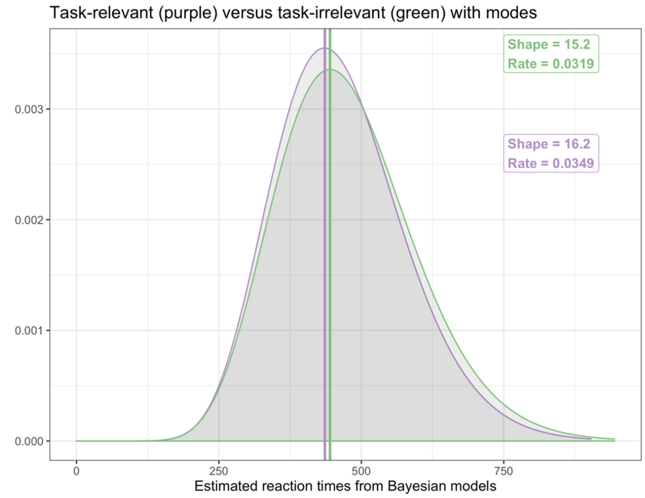
Priors. The prior for the difference in mode was modeled using an uninformative prior as a normal distribution with a mean of 0 and a standard deviation of 32 while the prior for the difference in log precision $\left(\frac{1}{\sqrt{s^2}}\right)$ was modeled as a normal distribution with a mean of 0 and a standard deviation of 3.2.

Main Effect of Task Relevance. For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The overall difference in mode between task-relevant and task-irrelevant trials was −9.91 ms (95% CI [−21.69 to 1.97]). The Rhat was 1.001, and the effective sample size was 14,000. The overall difference in log precision was 0.12 [−0.03 to 0.27]. The Rhat was 1.001, and the effective sample size was 15,000. See Figure 6 for posterior estimates of the gamma distributions following task-relevant and task-irrelevant trials. As is evident from the credible intervals, the estimates for both parameters overlap with zero and thus there is not convincing evidence for a difference in either central tendency or spread between task-relevant and task-irrelevant trials.

Task-Relevant Inner Speech Against All Other Trials. We tested task-relevant inner speech trials against all other trials. For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The overall difference in mode between task-relevant inner speech trials and the other types of trials was 22.97 ms (95% CI [10.46–35.81]). The Rhat was 1.001, and the effective sample size was 15,000. The overall difference in log precision was −0.42 [−0.59 to −0.24]. The Rhat was 1.001, and the effective sample size was 15,000. As is evident from the credible intervals, neither of the estimates for the parameters overlap with zero and thus there is convincing evidence for a difference in both central tendency and spread between task-relevant inner speech trials and all other types of trials. See Figure 7 for estimated gamma distributions on trials preceded by task-relevant inner speech and all other trials.

Figure 6

Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant (Purple) and Task-Irrelevant (Green)



Note. Vertical lines indicate modes. See the online article for the color version of this figure.

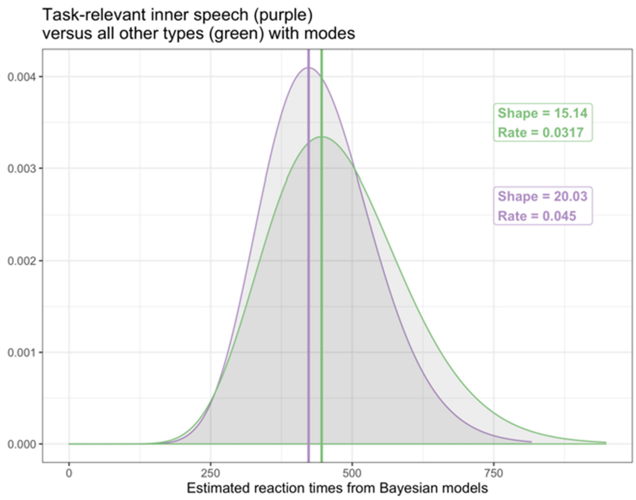
Task-Relevant Inner Speech Against Task-Relevant Non-Inner Speech. At the request of a reviewer, we tested trials preceded by a task-relevant inner speech against trials preceded by task-relevant experience not in the form of inner speech. This comparison was designed to check that the effect of the task-relevant inner speech was not driven by any main effect of task relevance. For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The comparison was not preregistered, and the model specification can be accessed through the online supplemental materials alongside the other models. The overall difference in mode between task-relevant inner speech trials and task-relevant non-inner speech trials was 16.82 ms (95% CI [2.73–30.61]). The Rhat was 1.001, and the effective sample size was 15,000. The overall difference in log precision was −0.39 [−0.58 to −0.19]. The Rhat was 1.001, and the effective sample size was 15,000. As is evident from the credible intervals, neither of the estimates for the parameters overlap with zero and thus there is convincing evidence for a difference in both central tendency and spread. See Figure 8.

Interim Summary

The results of our preregistered analyses suggest that task-relevant inner experience is associated with faster reaction times to infrequently occurring prompts. However, these results are in conflict with our unregistered hierarchical Bayesian models which suggest

Figure 7

Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant Inner Speech Trials (Purple) and All Other Types of Trials (Green)

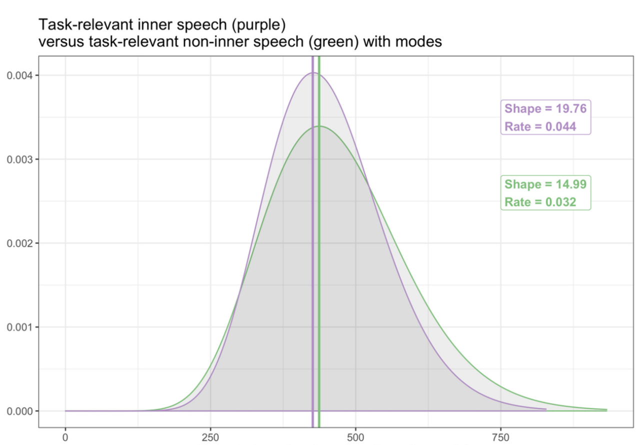


Note. Vertical lines indicate modes. See the online article for the color version of this figure.

that only trials preceded by task-relevant inner speech are associated with faster and less distributed reaction times compared with other types of experience. Inner speech occurred more frequently than the other four types of inner experience, contrasting with findings from Descriptive Experience Sampling studies where each type of experience occurs in 20%–25% of sampled moments (Heavey & Hurlburt, 2008). This may be a product of the experience types being mutually exclusive in our experiment while it is possible to report for example both "Inner voice" and "Feelings" in Descriptive Experience Sampling studies. On trials where participants reported inner speech, they were also asked to report how condensed, dialogic, and evaluative their inner voice was, and how much they had the experience of other people's voices. These reports were comparable with findings from questionnaire studies using similar questions (Alderson-Day et al., 2018).

Figure 8

Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant Inner Speech Trials (Purple) and Task-Relevant Non-Inner Speech Trials (Green)



Note. Vertical lines indicate modes. See the online article for the color version of this figure.

Because we conducted some additional analyses that we had not preregistered—notably the hierarchical Bayesian models—we decided to conduct the entire experiment again as a replication. The replication was also preregistered (https://osf.io/dvwbg).

Method: Experiment 2 (Replication)

The method was almost identical to the method in Experiment 1 with one change: Instead of eight trials per person, we had 12 trials per person to allow us to more robustly test the interaction between trial progression and inner experience. The order of the interstimulus intervals was again randomized for each participant (30 s twice, 40 s, 50 s twice, 60 s twice, 70 s, 120 s twice, 150 s, 180 s). Because this increased the duration of the experiment from approximately 20 min to approximately 26 min, we also increased the compensation from £3 to £4. The data were collected in January 2022 (Table 4).

Results: Experiment 2 (Replication)

Descriptive Statistics

Reaction Times

As detailed in the preregistration, we excluded trials with reaction times below 200 ms and trials from participants who missed three prompts (24 participants). Aside from these participants whose data were not included, we excluded 76 trials where other participants were too slow to respond (RT $>$ 1,000 ms, 2.9% of all trials). As predicted, the reaction time data were gamma rather than normally distributed (see Figure 9).

The distribution of response times was very similar to that observed in Experiment 1. Participants had a mean reaction time of 468.3 ms ($SD = 125.61$) and a median reaction time of 442.95 ms, supporting the assumption that reaction times were positively skewed and thus followed a gamma rather than a normal distribution.

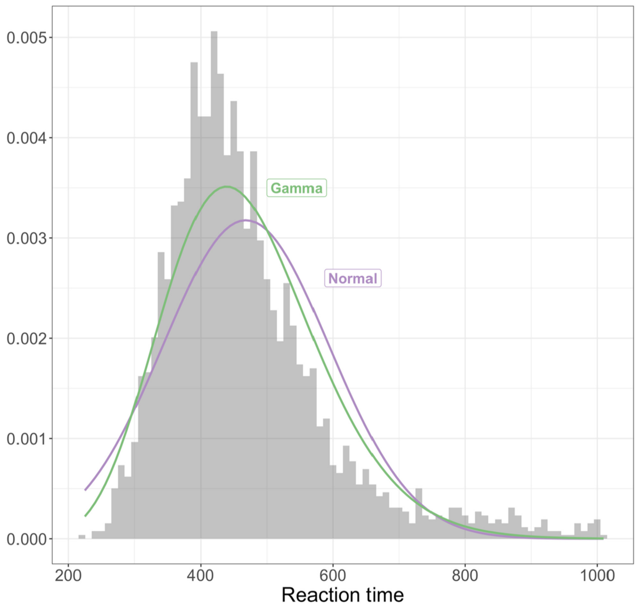
The fitted models' log-likelihoods are as follows: Normal $= -16{,}180.03$; gamma $= -15{,}934.14$. The higher the log-likelihood, the better the fit, confirming that the gamma distribution fits the data best. For that reason and for theoretical reasons (see Lo & Andrews, 2015), we use a gamma distribution for the remaining analyses.

Table 4

Demographic Data for All Participants in the Replication

N (after exclusions)	222
Excluded (failed to respond to three prompts)	24
Female	134
Male	88
Median age (range)	34 (18–76)
Median time spent (range)	23 min 24 s (18 min 30 s–69 min 59 s)

Figure 9
Fitted Gamma and Normal Distributions to All Trials Across Participants



Note. The gamma distribution models a continuous distribution with two parameters (shape and rate) which is often used to model wait times and other phenomena that are always positive and skewed. When the shape parameter is >1, the distribution is positively skewed. The normal distribution is symmetric and models a continuous distribution with two parameters (mean and standard deviation). See the online article for the color version of this figure.

Experience Questions

See Table 5 and Figure 10 for proportions of reported inner experience types. The distribution resembled the one observed in Experiment 1. Chi-square tests suggested that task relevance and experience type were not independent when excluding trials where participants responded "I don't know" to the task relevance question ($\chi^2 = 112.45$, $df = 4$, $p < .001$).

Inner Speech Questions

See Figure 11 for how participants answered the specific questions about the nature of their inner speech as well as how our experiment compares with a previous study using similar items (Alderson-Day et al., 2018). Despite the differences in Likert scales (five-point in ours and seven-point in the original VISQ-R study), it is evident that proportions are comparable. In our sample, 26.1% of participants never reported experiencing inner speech (comparable to 24.1% in Experiment 1).

When asked whether they had talked to themselves to stay focused during the experiment, 165 participants (74.3%) said that they had and 57 participants (25.7%) said that they had not. This is comparable to proportions in Experiment 1.

Hierarchical Bayesian Model: Replication

For all models, the full model specifications can be found in the online supplemental materials. The different effects of interest (main effect of task relevance and the interaction between task relevance and inner speech) were tested by defining different contrasts. See Figure 12 for a density plot of reaction times following task-relevant inner speech, task-irrelevant inner speech, task-relevant non-inner speech, and task-irrelevant non-inner speech.

Priors

We used uninformative priors for both the difference in mode and standard deviation. The prior for the difference in mode was modeled as a normal distribution with a mean of 0 and a standard deviation of 32 while the prior for the difference in log precision $\left(\frac{1}{\sqrt{s^2}}\right)$ was modeled as a normal distribution with a mean of 0 and a standard deviation of 3.2.

Main Effect of Task Relevance

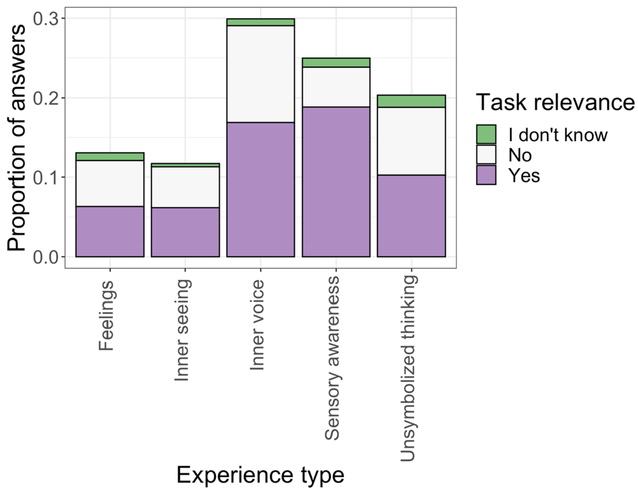
For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The overall difference in mode between task-relevant and task-irrelevant trials was 18.3 ms (95% CI [8.29–27.84]). The Rhat was 1.001, and the effective sample size was 14,000. The overall difference in log precision was −0.28 [−0.4 to −0.16]. The Rhat was 1.001, and the effective sample size was 15,000. See Figure 13 for posterior estimates of the gamma distributions following task-relevant and task-irrelevant trials. As is evident from the credible interval, neither of the estimates for the parameters overlap with zero and thus there is convincing evidence for a difference in both central tendency or spread between task-relevant and task-irrelevant trials. To better compare with Experiment 1, we also conducted the Bayesian models on only the first eight trials from the replication data. Here, the overall difference in mode between task-relevant and task-irrelevant trials was 12.60 ms [0.77–23.98], and the overall difference in log precision was −0.25 [−0.4 to −0.10].

Table 5
Reported Types of Inner Experience in Percentages Across All Prompts (Twelve Per Participant)

Experience type	Task-relevant (count)	Task-irrelevant (count)	"I don't know" responses (count)	Percentage of total responses
Feelings	169	153	26	13.06%
Inner seeing	165	136	11	11.71%
Inner voice	450	324	23	29.92%
Sensory awareness	502	134	29	24.96%
Unsymbolized thinking	273	228	41	20.35%

Figure 10

Visualization of Reported Types of Inner Experience and Whether They Were Relevant to the Task or Not Across All Prompts (Twelve Per Participant)



Note. See the online article for the color version of this figure.

Task-Relevant Inner Speech Against All Other Trials

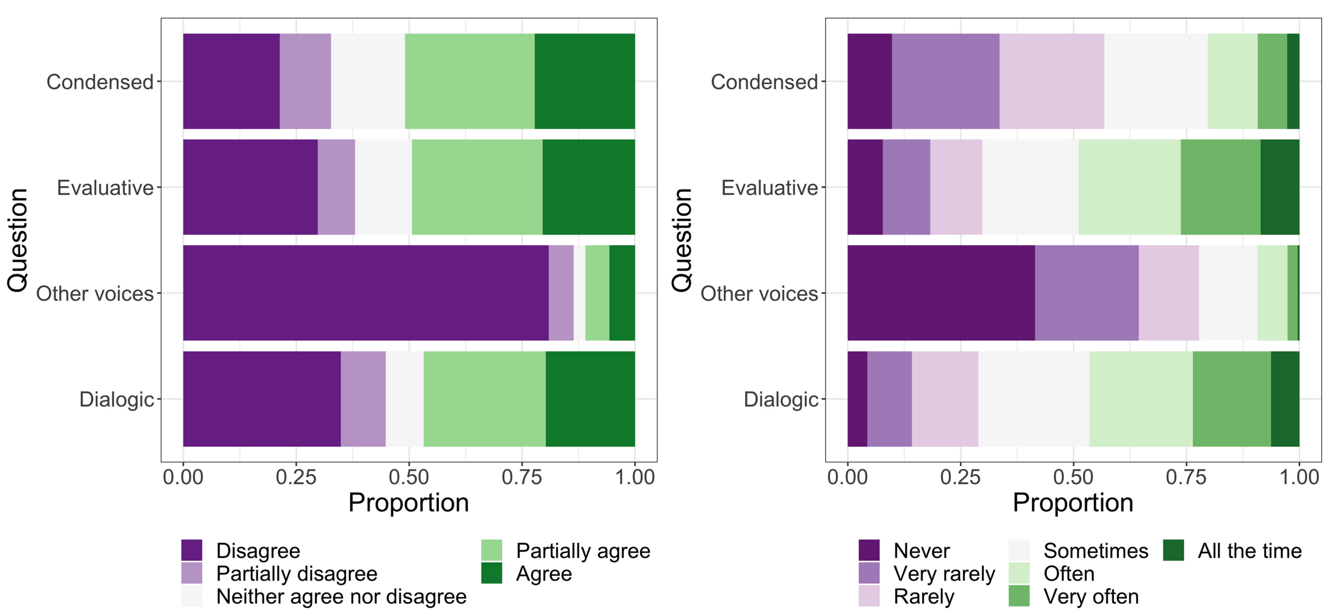
For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The overall difference in mode between task-relevant inner speech trials and all other trials was 18.08 ms (95% CI [7.51–28.57]). The Rhat was 1.001, and the effective sample size was 15,000. The overall difference in log precision was −0.3 [−0.45 to −0.15]. The Rhat was 1.001, and the effective sample size was 3,300. As is evident from the credible intervals, neither of the estimates for the parameters overlap with zero and thus there is convincing evidence for a difference in both central tendency or spread between task-relevant inner speech trials and other types of trials. See Figure 14 for estimated gamma distributions on trials preceded by task-relevant inner speech and all other trials.

Task-Relevant Inner Speech Against Task-Relevant Non-Inner Speech

At the request of a reviewer, we tested trials preceded by task-relevant inner speech against trials preceded by task-relevant experience not in the form of inner speech. This comparison was designed to check that the effect of task-relevant inner speech was not driven by any main effect of task relevance. For this model, we had three chains and 10,000 iterations (first 5,000 discarded). The overall difference in mode between task-relevant inner speech trials and all other trials was 9.82 ms (95% CI [−1.47 to 21.14]). The Rhat was 1.002, and the effective sample size was 3,100. The overall difference in log precision was −0.21 [−0.37 to −0.05]. The Rhat was 1.001, and the effective sample size was 4,300. As is evident from the credible intervals, the difference in modes overlaps with zero so there is only weak evidence for a difference in the central tendency in reaction time. The credible interval of the difference in log precision, on the other hand, does not overlap with zero, and thus there *is* convincing evidence that task-relevant inner speech trials show lower variance. See Figure 15 for estimated gamma distributions on trials preceded by task-relevant inner speech and task-relevant non-inner speech trials.

Figure 11

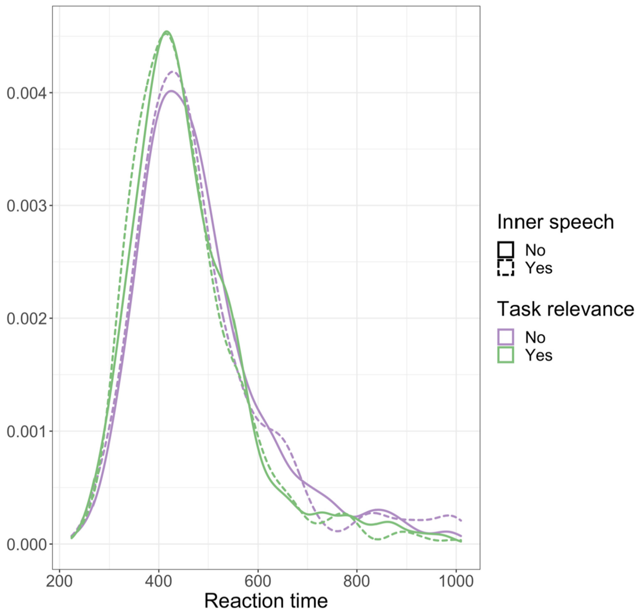
On the Left, We See the Answers to the Inner Speech Questions in the Present Experiment



Note. On the right, we see the answers from VISQ-R (Alderson-Day et al., 2018). Note that we had a five-point Likert scale while Alderson-Day et al. had a seven-point Likert scale. See the online article for the color version of this figure.

Figure 12

Visualization of the Densities of Reaction Times in the Four Combinations of Task Relevance and Inner Speech



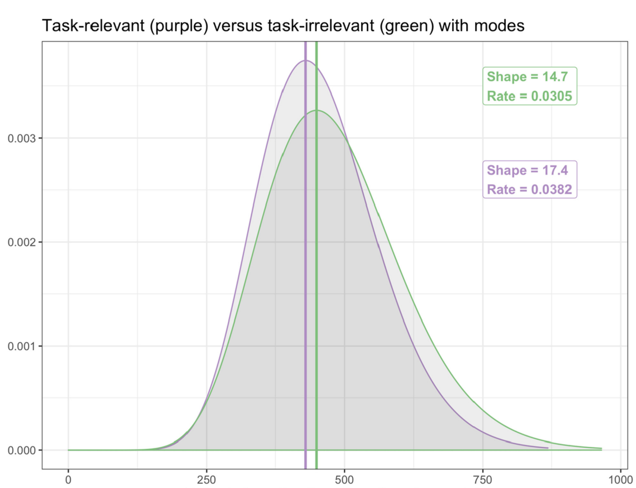
Note. As in the original experiment, we used hierarchical Bayesian modeling to investigate the distributions. See the online article for the color version of this figure.

Discussion

Across two online experiments (an original and a near-identical replication), we found evidence that participants can respond faster and with less variable response times to infrequently occurring prompts by talking to themselves about the task. Task-relevant inner experience in general was associated with faster reaction times (especially in the replication which was longer and thus more demanding of self-control), and this interacted with inner speech. These findings suggest two important things: (a) Humans can use their inner voice for focused control of attention in tedious situations, (b) This use of inner voice can decrease reaction times and reduce variability in responses.

Figure 13

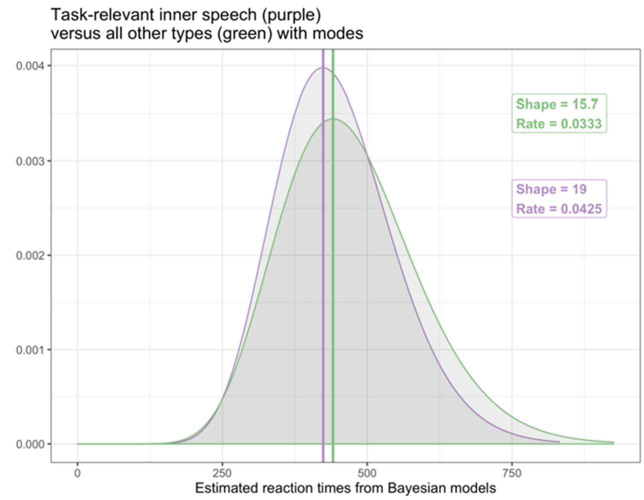
Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant Trials (Purple) and All Task-Irrelevant Trials (Green)



Note. Vertical lines indicate modes. See the online article for the color version of this figure.

Figure 14

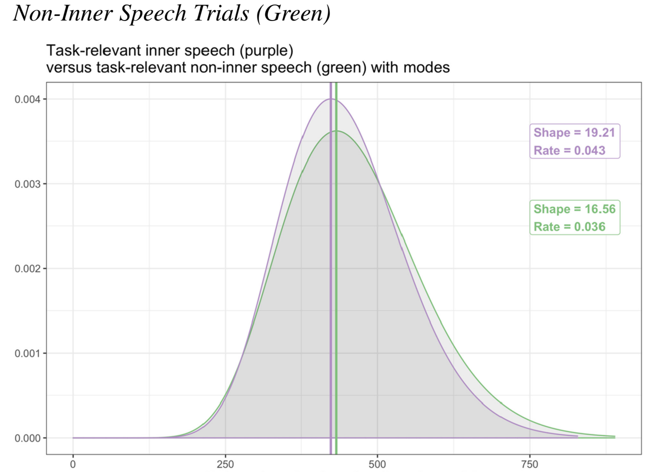
Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant Inner Speech Trials (Purple) and All Other Types of Trials (Green)



Note. Vertical lines indicate modes. See the online article for the color version of this figure.

Figure 15

Posterior Estimates of the Reaction Time Gamma Distributions on Task-Relevant Inner Speech Trials (Purple) and Task-Relevant Non-Inner Speech Trials (Green)



Note. Vertical lines indicate modes. See the online article for the color version of this figure.

Interaction Between Task Relevance and Inner Speech

Traditional generalized linear mixed models were unable to detect the interaction effect between task relevance and inner speech because the differences appeared to be in both the central tendency and the variability of the reaction time distribution, and not just the central tendency. To further explore these differences, we applied hierarchical Bayesian models to the data and found good evidence for the expected interaction effect. Because the hierarchical Bayesian analysis was not included in our preregistration and to test the robustness of our results, we decided to run a near-identical replication which confirmed the pattern found in the first experiment. One important difference was that the main effect of task relevance was not supported by the hierarchical Bayesian models in the original experiment. However, in the replication, which featured 12 trials per person instead of eight, there appeared to be robust evidence for the effect of task relevance in the Bayesian model. This may suggest that task relevance becomes more important as the self-control demands rise with time. This is consistent with previous research on sustained attention (Lichstein et al., 2000), and the relationship with inner speech would be an interesting avenue for future studies to explore both empirically and through formal theoretical models. A direct comparison between task-relevant inner speech trials and task-relevant non-inner speech trials in the replication experiment suggested that the contribution of inner speech over and above the effect of task relevance lay mostly in reducing the *variability* in responses and not so much in making the responses *faster*.

Modeling of Reaction Times

We decided to use gamma distributions to model the reaction times in this study because gamma distributions are associated with wait times and serial stages of processing of events that must occur before a given response. Each of these stages has a finish time that is exponentially distributed (Van Zandt & Ratcliff, 1995). In the present case, the stages conceivably are as follows:

1. Visually registering the circle prompt.
2. Returning attention from being off-task (if necessary).
3. Recalling and preparing the appropriate reaction (pressing the button quickly).
4. Executing the motor, n.d.

Task relevance presumably reduces reaction time by skipping stage 2 (removing one exponentially distributed component). The reason why task-relevant inner speech would be associated with a narrower distribution is conceivable that you can use your inner voice to prime the appropriate reaction (i.e. enhancing attention to the task and its requirements at stage 3). This fits well with recent findings from sport psychology suggesting that a self-talk intervention can increase attentional control (Galanis et al., 2022). It is easier to imagine the many stages involved in responding to a simple stimulus such as ours by imaging what would happen if a computer were mechanically programmed to perform this task. If such a computer had to react to this stimulus, its reaction times would be very fast and would likely produce an exponential distribution centered around the clock speed. Because humans must balance other tasks and attend to other details in their perceptual environment, the change in visual stimulus requires an additional step of aligning attention with the task at hand. This additional step is required unless endogenous control mechanisms prevent attention from drifting to other matters. Task-directed inner speech may be one mechanism, or corollary, of this endogenous control process. The distinctions outlined here between human and computer attention are closely related to Posner's ideas of attention being divided into alerting, orienting, and executive control (Posner, 2016).

For the wider literature on inner speech and its role in cognition, it is important to note that inner voice only had a beneficial effect on reaction times if it was also task-relevant. Inner speech is not simply across-the-board beneficial which emphasizes two important things: (a) Inner speech is a tool that can be used more or less productively; (b) The content of inner speech makes a difference for behavior and is not just an epiphenomenon of consciousness. However, the beneficial effects of task-relevant inner speech could depend on the specific task as there are other examples in the mind-wandering literature of task-related thoughts (usually negative) actually interfering with task performance (Gonçalves et al., 2017; Maillet & Rajah, 2013; McVay & Kane, 2009). These tasks are usually more continuous (e.g. the Sustained Attention to Response Task or the Metronome Response Task)—in our case, we believe that any thoughts related to the task would prime the reaction time response and thus be beneficial.

Limitations of the Current Study

The main limitation of the current study is the reliance on self-reported data collected online. Although jsPsych has performed well on benchmark tests of reaction time (de Leeuw & Motz, 2016), the differences reported in the current paper are very fine-grained so even small amounts of noise due to software or hardware differences could distort the data. The fact that our results replicate, however, indicates that incidental noise did not create the results. We cannot think of any reasons why any noise components related to the experimental setup would affect task-relevant inner speech trials differently than other trial types.

It could also be problematic that we ask participants *after* they have responded to a button press what they were experiencing immediately *before* the prompt and then use those answers to predict the reaction times. The alternative was to alternate randomly between experience prompts and reaction time prompts but that would cause different problems—first, the experiment would be longer because we would have to insert twice as many wait times to allow participants to go off-task again, and second, we would then be even less confident that the experience reported after the experience prompt had an effect on the reaction time. Another issue with the procedure is that participants might have reported inner experience that "fitted" their reaction time, i.e. if they felt they had been fast, they would answer "task-relevant" and if they felt they had been slow, they would answer "task-irrelevant." However, we do not believe participants would have been able to detect such small differences in reaction times when trials occurred several minutes apart.

Even though participants were instructed to keep their gaze fixed on the fixation cross during wait times, it is possible that they deviated from these instructions and paid attention to their phone or a book instead. While the online setup necessarily decreased experimental control, we do not believe that the participants' potential distraction necessarily invalidates our results. If they were looking at their phone or reading a book while waiting, they would presumably be slower to respond and report that they were off task. A reviewer brought up the concern that participants may not have understood what the different experience type categories meant (e.g. "unsymbolized thinking"). In experiments conducted online, it is of course difficult to ensure that participants fully understand the task instructions but we did not get any feedback from participants in the free answer blocks to indicate that they were confused. Additionally, all the experience types were chosen at least a few times which they presumably would not have been if participants did not understand what they meant.

Conclusion

Investigating the influence of inner speech on behavior is a challenging pursuit, mainly because inner speech itself is an elusive concept. We here explored a new method combining experience sampling and attentional control and found that people to a large extent talk to themselves to stay focused on a boring task. We also found that this task-relevant inner speech was associated with reaction times that were not only faster but also less distributed than task-relevant non-inner speech, task-irrelevant non-inner speech, and task-irrelevant inner speech. Our findings across two experiments suggest that inner speech can be recruited as a tool for attentional and behavioral control.

References

Aitchison, C., Turner, L. A., Ansley, L., Thompson, K. G., Micklewright, D., & Gibson, A. S. C. (2013). Inner dialogue and its relationship to perceived exertion during different running intensities. *Perceptual and Motor Skills*, *117*(1), 11–30. https://doi.org/10.2466/06.30.PMS.117x11z3

Alderson-Day, B., & Fernyhough, C. (2015a). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, *6*, Article 517. https://doi.org/10.3389/fpsyg.2015.00517

Alderson-Day, B., & Fernyhough, C. (2015b). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

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. https://doi.org/10.1016/j.concog.2018.07.001

De Leeuw, J. R. (2015). Jspsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

Dickens, Y. L., Van Raalte, J., & Hurlburt, R. T. (2018). On investigating self-talk: A descriptive experience sampling study of inner experience during golf performance. *The Sport Psychologist*, *32*(1), 66–73. https://doi.org/10.1123/tsp.2016-0073

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. https://doi.org/10.1016/S0749-596X(02)00511-9

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Galanis, E., Hatzigeorgiadis, A., Comoutos, N., Papaioannou, A., Morres, I. D., & Theodorakis, Y. (2022). Effects of a strategic self-talk intervention on attention functions. *International Journal of Sport and Exercise Psychology*, *20*(5), 1368–1382. https://doi.org/10.1080/1612197X.2021.1963304

Gonçalves, Ó. F., Rêgo, G., Oliveira-Silva, P., Leite, J., Carvalho, S., Fregni, F., Amaro Jr E., & Boggio, P. S. (2017). Mind wandering and the attention network system. *Acta Psychologica*, *172*, 49–54. https://doi.org/10.1016/j.actpsy.2016.11.008

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hartshorne, J. K. (2020). How massive online experiments (MOEs) can illuminate critical and sensitive periods in development. *Current Opinion in Behavioral Sciences*, *36*, 135–143. https://doi.org/10.1016/j.cobeha.2020.09.005

Hartshorne, J. K., de Leeuw, J. R., Goodman, N. D., Jennings, M., & O'Donnell, T. J. (2019). A thousand studies for the price of one: Accelerating psychological science with Pushkin. *Behavior Research Methods*, *51*(4), 1782–1803. https://doi.org/10.3758/s13428-018-1155-z

Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., & Theodorakis, Y. (2011). Self-talk and sports performance: A meta-analysis. *Perspectives on Psychological Science*, *6*(4), 348–356. https://doi.org/10.1177/1745691611413136

Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, *17*(3), 798–810. https://doi.org/10.1016/j.concog.2007.12.006

Hurlburt, R. T., Heavey, C. L., Lapping-Carr, L., Krumm, A. E., Moynihan, S. A., Kaneshiro, C., Brouwers, V. P., Turner, D. K., & Kelsey, J. M. (2022). Measuring the frequency of inner-experience characteristics. *Perspectives on Psychological Science*, *17*(2), 559–571. https://doi.org/10.1177/1745691621990379

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, *10*(11), 480–486. https://doi.org/10.1016/j.tics.2006.09.002

Kray, J., Eber, J., & Karbach, J. (2008). Verbal self-instructions in task switching: A compensatory tool for action-control deficits in childhood and old age? *Developmental Science*, *11*(2), 223–236. https://doi.org/10.1111/j.1467-7687.2008.00673.x

Lichstein, K. L., Riedel, B. W., & Richman, S. L. (2000). The Mackworth Clock Test: A computerized version. *The Journal of Psychology*, *134*(2), 153–161. https://doi.org/10.1080/00223980009600858

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, Article 1171. https://doi.org/10.3389/fpsyg.2015.01171

Maillet, D., & Rajah, M. N. (2013). Age-related changes in frequency of mind-wandering and task-related interferences during memory encoding and their impact on retrieval. *Memory*, *21*(7), 818–831. https://doi.org/10.1080/09658211.2012.761714

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10.1037/a0014104

Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*(3), 376–382. https://doi.org/10.1002/acp.3404

Morin, A., Uttl, B., & Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Procedia—Social and Behavioral Sciences*, *30*, 1714–1718. https://doi.org/10.1016/j.sbspro.2011.10.331

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *141*(4), 788–798. https://doi.org/10.1037/a0027968

Nedergaard, J. S. K., Christensen, M. S., & Wallentin, M. (2022). *Mind over body: Interfering with the inner voice is detrimental to endurance performance*. https://doi.org/10.31234/osf.io/bfj9w

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*, Article 103102. https://doi.org/10.1016/j.concog.2021.103102

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*. https://doi.org/10.3758/s13423-022-02144-7

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, No. 125.10, pp. 1–10).

Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, *69*(10), 1864–1875. https://doi.org/10.1080/17470218.2014.937446

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

Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*, Article 430. https://doi.org/10.3389/fpsyg.2013.00430

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 1–5. https://doi.org/10.1037/a0030954

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review*, *14*(3), 527–533. https://doi.org/10.3758/BF03194102

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Su, Y.-S., & Masanao, Y. (2021). *R2jags: Using R to Run "JAGS."* https://CRAN.R-project.org/package=R2jags

Thibodeaux, J., & Winsler, A. (2018). What do youth tennis athletes say to themselves? Observed and self-reported self-talk on the court. *Psychology of Sport and Exercise*, *38*, 126–136. https://doi.org/10.1016/j.psychsport.2018.06.006

Tod, D., Hardy, J., & Oliver, E. (2011). Effects of self-talk: A systematic review. *Journal of Sport and Exercise Psychology*, *33*(5), 666–687. https://doi.org/10.1123/jsep.33.5.666

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, *135*(2), 252–256. https://doi.org/10.1016/j.actpsy.2010.07.008

Uttl, B., Morin, A., Faulds, T., Hall, T., & Wilson, J. (2012). Sampling inner speech using text messaging. *Proceedings of the Canadian Society for Brain, Behavior, and Cognitive Science, Kingston, ON, Canada*. (pp. 287–290). https://www.researchgate.net/profile/Alain-Morin-2/publication/281131203_Sampling_inner_speech_using_text_messaging/links/5d2907c7299bf1547cb11e34/Sampling-inner-speech-using-text-messaging.pdfLocation

Uttl, B., Morin, A., & Hamper, B. (2011). Are inner speech self-report questionnaires reliable and valid? *Procedia—Social and Behavioral Sciences*, *30*, 1719–1723. https://doi.org/10.1016/j.sbspro.2011.10.332

Van Raalte, J. L., Cornelius, A. E., Brewer, B. W., & Hatten, S. J. (2000). The antecedents and consequences of self-tal0k in competitive tennis. *Journal of Sport and Exercise Psychology*, *22*(4), 345–356. https://doi.org/10.1123/jsep.22.4.345

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, *2*(1), 20–54. https://doi.org/10.3758/BF03214411

Welhaf, M. S., Smeekens, B. A., Meier, M. E., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2020). The worst performance rule, or the not-best performance rule? Latent-variable analyses of working memory capacity, mind-wandering propensity, and reaction time. *Journal of Intelligence*, *8*(2), Article 25. https://doi.org/10.3390/jintelligence8020025

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Zourbanos, N., Hatzigeorgiadis, A., Bardas, D., & Theodorakis, Y. (2013). The effects of self-talk on dominant and nondominant arm performance on a handball task in primary physical education students. *The Sport Psychologist*, *27*(2), 171–176. https://doi.org/10.1123/tsp.27.2.171

Received April 15, 2022
Revision received December 21, 2022
Accepted December 23, 2022 ■

Article IV

Not everyone has an inner voice: Behavioral consequences of anendophasia

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

It is commonly assumed that inner speech – the experience of thought as occurring in a natural language – is both universal and ubiquitous. Recent evidence, however, suggests that similar to other phenomenal experiences like visual imagery, the experience of inner speech varies between people, ranging from constant to non-existent. We propose a name for a lack of the experience of inner speech – anendophasia – and report four experiments examining some of its behavioral consequences. We found that people who report low levels of inner speech have lower performance on a verbal working memory task and have more difficulty performing rhyme judgments based on images. Task switching performance, previously linked to endogenous verbal cueing, was unaffected by differences in inner speech, as was a visual discrimination task. We also report results of a questionnaire showing anendophasia to be associated with a range of experiential differences ranging from experiencing earworms to memory for conversations. We discuss our findings in relation to aphantasia, condensed versus expanded inner speech, and unsymbolized thinking.

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

Word count: 7167

1 Introduction

It is frequently claimed that everyone has an inner voice, and that most of our waking hours are filled with internal monologue (e.g., 'We all hear a voice inside our brain, commonly called "inner voice", "inner speech" or referred to as "verbal thoughts" '; Perrone-Bertolotti et al., 2014, p. 221). Recent evidence – both anecdotal accounts and more systematic investigations – challenge this view. In mass media, 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). Systematic investigations have focused on auditory imagery as a proxy for inner speech (Dawes, Keogh, Andrillon, & Pearson, 2020; Hinwar & Lambert, 2021) and found that auditory imagery, like visual imagery, varies from entirely absent to ubiquitous across individuals.

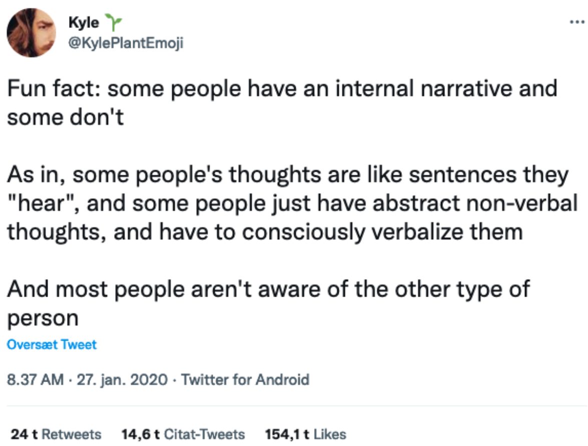


Figure 1. Viral tweet from @KylePlantEmoji about the presence or absence of inner speech. Screenshot from November 17th 2022.

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. We explore this intriguing possibility in the present study. If there *are* differences, this helps us understand the extent to which people's cognition may be differentially guided by language. If there are *no* differences, this could either mean that the measure of inner speech is invalid (i.e., people reporting more or less inner speech in fact have similar experiences but interpret questions concerning inner speech differently), or that differences in inner speech have no bearing on the behavioral measures in question, or that people who experience no inner speech do not differ in accuracy or speed because they rely on different processes or strategies than inner speech.

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 includes claims that inner speech constitutes (at least some types of) thought (Bermúdez, 2007; Carruthers, 2002; Clark, 1998; Frankish, 2018; Gauker, 2011; Morin, 2018) and theories that inner speech is crucially involved in behavioral control (Alderson-Day & Fernyhough, 2015; Cragg & Nation, 2010; Emerson & Miyake, 2003; Morin, Duhnych, & Racy, 2018).

1.1 Parallels with condensed inner speech and unsymbolized thinking

What do people mean when they say they do not experience inner speech? Anecdotally, some report on internet fora 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 not clear. Beyond informal self-reports, the existence of such non-verbal and non-perceptual 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 their inner experience, their reports are often consistent with what Hurlburt and colleagues have 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, 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). 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. Instead, the question is a single, undifferentiated whole. It is possible that unsymbolized thinking is continuous with inner speech with weak or absent conscious imagery since it to some extent appears to have similar semantic and syntactic structures as language (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.

Descriptive Experience Sampling has yielded another finding that is potentially relevant for what is experienced as a lack of inner speech: "wordless" inner speech (Hurlburt, Heavey, & Kelsey, 2013), akin to a series of tip of the tongue states. In such episodes, people often report experiencing the pace, rhythm, and linear sequence of speech without the experience of hearing or speaking words. The idea that inner speech may vary in how closely tied it is to audition and articulation plays an important role in several different conceptualizations of inner speech (Fernyhough, 2004; Grandchamp et al., 2019; Oppenheim & Dell, 2010). For example, the developmental psychologist Vygotsky thought that adult inner speech is an internalized form of children's overt speech, and that inner speech during this internalization is transformed to be more condensed in terms of both form and meaning. Vygotsky thought that the most condensed form of inner speech could be thought of as 'thinking in pure meanings' and thus be abstracted away from both phonological and articulatory specification (Vygotsky, 1962). Some theorists have suggested that the degree to which inner speech is experienced as condensed or expanded varies across both individuals and situations (Fernyhough, 2004; Grandchamp et al., 2019), for example under higher cognitive demands or social isolation (Brinthaupt, 2019).

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 with aphantasia had difficulties with autobiographical memory (Zeman et al., 2010). Generally, participants with aphantasia report weak or non-existing ability to visualize "in the mind's eye" (Dawes et al., 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 use 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.

1.3 The present study

Taking inspiration from aphantasia research on visual working memory, we can also test the **verbal working memory** performance of people reporting little to no inner speech. This allows us to both test whether verbal working memory and reported inner speech use are related and explore the possible compensatory strategies. In particular, we might expect difficulties with verbal working memory tasks requiring a high degree of phonological precision (Jacobs et al., 2018). In the present study, 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. To the extent that phonological similarity makes recall more difficult (Baddeley, 1966; Murray, 1968), less inner speech may be associated with a reduced phonological similarity effect.

To further probe participants' 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. We reasoned that although participants reporting low levels of inner speech would have no trouble naming the objects, less reliance on inner speech would make it harder to compare the names in memory – necessary for making a rhyme judgment.

There is robust evidence that inner speech is often recruited for behavioral control in **task switching** paradigms where participants have to switch between different task rules (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Goschke, 2000; 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 the verbal interference literature). We reasoned that people who do not habitually use inner speech might be selectively impaired when they have to rely on self-generated cues. On the other hand, it is possible that they have learned to rely on other strategies in which case no difference would be found.

There is considerable evidence that language induces more **categorical representations** from basic perception onward (e.g. Forder & Lupyan, 2019; Perry & Lupyan, 2014; 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 (Lupyan, 2012), may be reduced in people with less inner speech.

Thus, in the present study, we explore individual differences related to reported inner speech in four behavioral tasks: verbal working memory, rhyme judgment, task switching, and categorical and perceptual visual discrimination.

2 Methods

2.1 Participants

We recruited participants who had previously completed the Internal Representations Questionnaire (Roebuck & Lupyan, 2020) as part of unrelated studies, contacting participants with verbal factor scores < 3.5 (bottom 16%-ile) or > 4.25 (top 40%-ile) on the Verbal factor of the questionnaire which is largely centered on propensity to experience and rely on inner speech. The percentiles were asymmetrical because it was more difficult to recruit participants reporting low levels of inner speech, and because the distribution in verbal scores on the IRQ is negatively skewed. Recruiting for example the top and bottom quartiles instead would have resulted in a "low inner speech" group who did not in fact have very low verbal representation scores. We received ethical approval from the University of Wisconsin-Madison. Ten participants 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 low verbal factor scores. The two groups were balanced in terms of age, gender, education level, dyslexia, and first language. See Table 1. Because of a technical error, demographic data is missing for one participant with less inner speech.

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	Median = 37; range = 18-67	Median = 39; range = 18-70	t(88.43) = -0.19; p = .849
Gender	22 female, 25 male	19 female, 26 male	$\chi^2(1) = 0.05$; p = .816
Native English-speaker	47 native speakers, 0 non-native speakers	41 native speakers, 4 non-native speakers	$\chi^2(1) = 2.49$; p = .114
Dyslexia	46 non-dyslexic, 1 self-diagnosed	44 non-dyslexic, 1 self-diagnosed	$\chi^2(1) < 0.01$; p = 1
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 = .815

2.2 Method: Verbal working memory

2.2.1 Materials and procedure. We used word sets from Baddeley (1966) because they were designed to be equivalent in other respects than phonological and orthographical similarity. 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”). On a given trial, participants saw five words in random order from one of the sets presented sequentially in writing and were then asked to type them back in the right order. 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 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. See Figure 2.

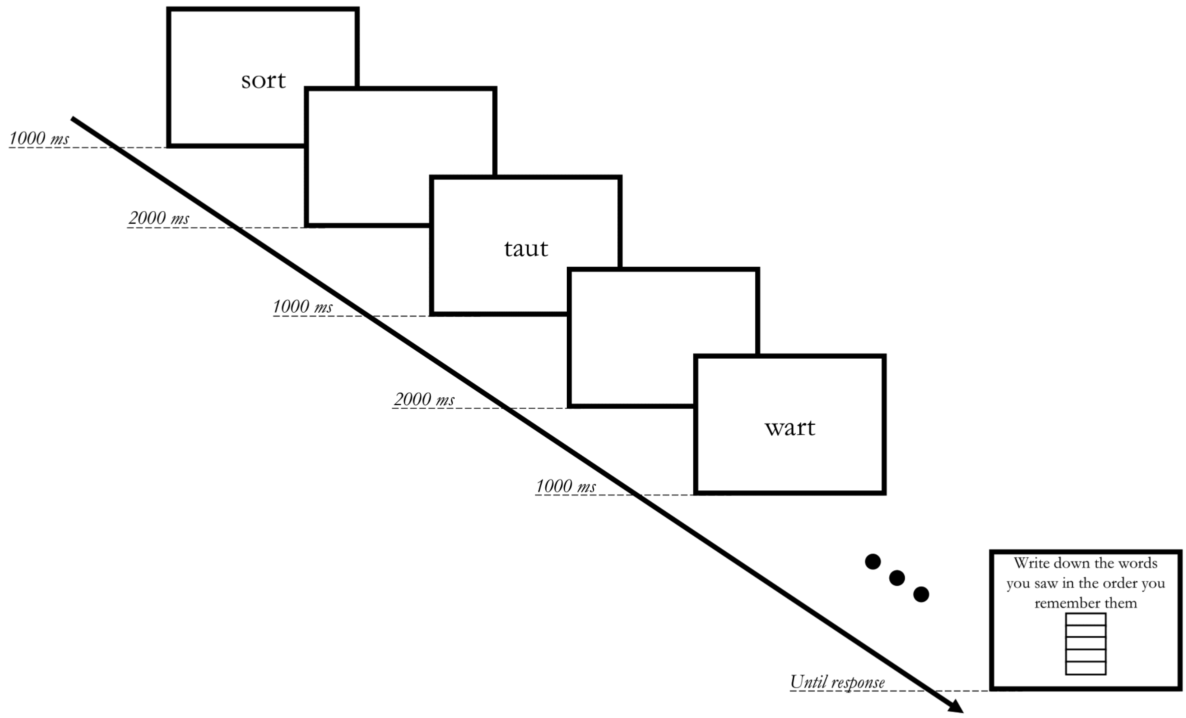


Figure 2. A sketch of the procedure in the verbal working memory experiment. In this example, the words are drawn from the phonological similarity set. Participants saw five words on each trial - three words are presented on the figure for ease of interpretation.

2.3 Method: Rhyme judgments

2.3.1 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 Appendix A 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 and represented the words we selected for the rhyme pairs. 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 orthographic rhymes, 20 non-orthographic rhymes, and 20 no-rhyme control trials). See Figure 3. Nameability scores for the images were collected from a separate set of 20 participants who were asked to label all the images. The nameability scores represent the proportion of participants who provided the target label.

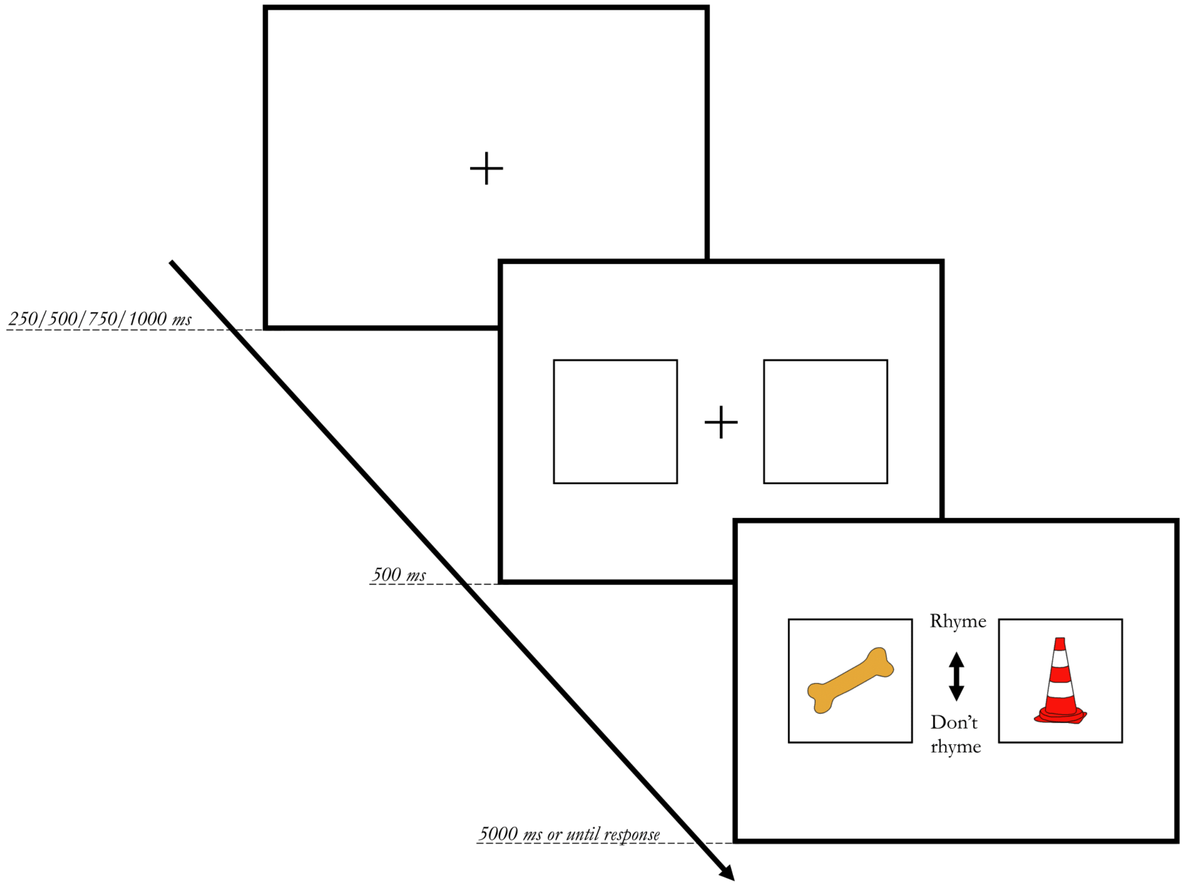


Figure 3. A sketch of a rhyme judgment trial. The stimuli here exemplify an orthographic rhyme – "bone" and "cone" – and the correct answer would therefore be "Rhyme".

2.4 Method: Task switching

2.4.1 Materials 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. 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. 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.

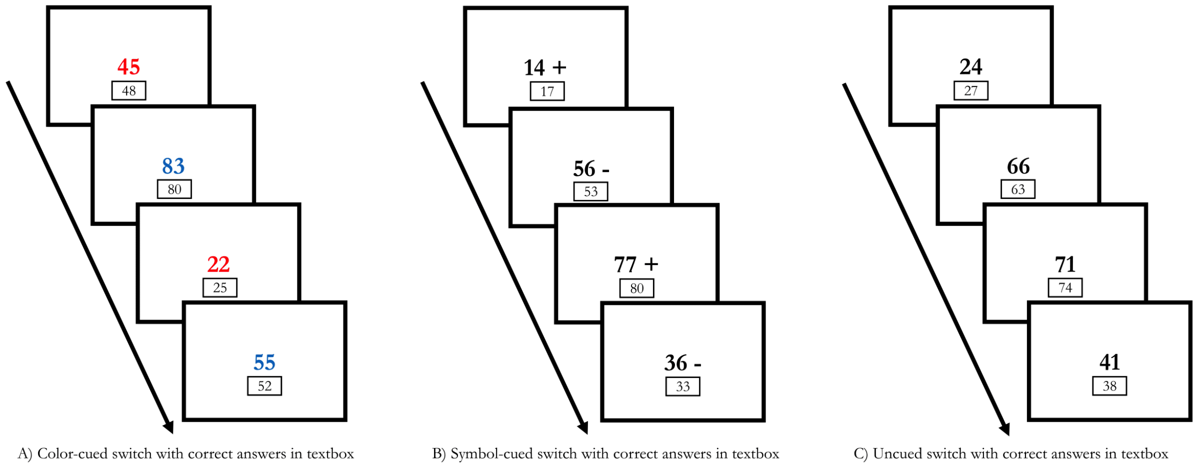


Figure 4. A sketch of the three switched conditions in the task switching experiment. Figure A shows four color-cued switch trials with correct answers, Figure B shows four symbol-cued switch trials with correct answers, and Figure C shows four un-cued switch trials with correct answers.

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.

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 visual 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").

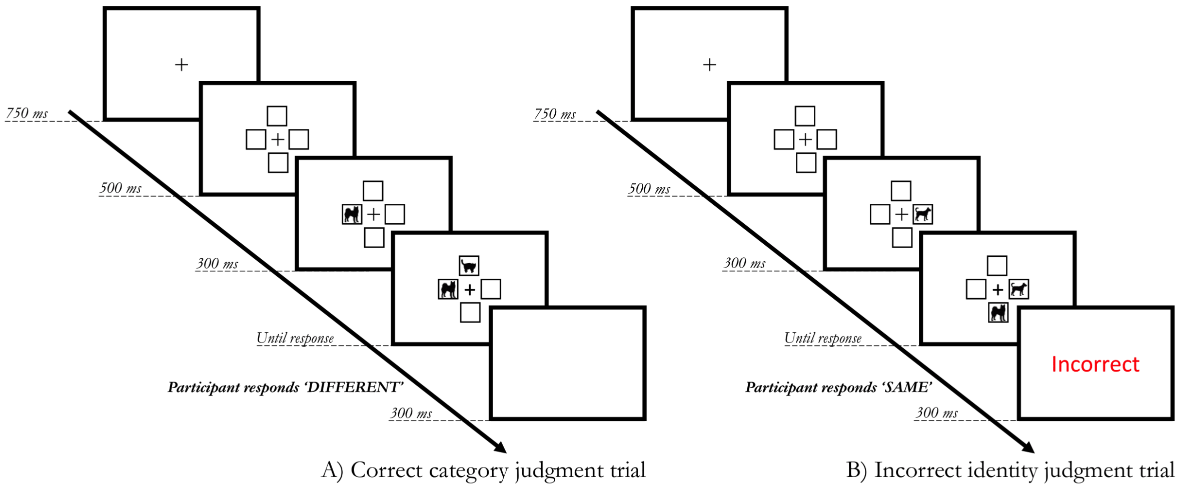


Figure 6. A sketch of the two conditions of the category judgment experiment. On Figure A, we see a correct category judgment trial where the participant responds that the cat and dog silhouettes represent different animals. On Figure B, we see an incorrect identity judgment trial where the participant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

2.6 Method: Questionnaire

After completing the four experiments, participants answered custom 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, Mitrenga, Wilkinson, McCarthy-Jones, & Fernyhough, 2018). See Appendix B for the full set of custom questions.

2.7 Data analysis

All analyses were conducted in R version 4.1.3 (see Appendix C for packages and citations). 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 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.

3 Results

3.1 Verbal working memory

3.1.1 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 7).

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

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)
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	0.10
More inner speech	Phonological similarity set	3.43	0.16	4.11	0.10
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

3.1.2 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) which were in turn remembered worse than the dissimilar words (M = 3.94) (β = -0.33; SE = 0.08; t = -3.98; p < .001). 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). This effect remained significant if we disregarded the order in which participants responded, counting only whether they recalled the correct words (β = 0.19; SE = 0.08; t = 2.57; p = .012). There were no interaction effects (all p > .104), although numerically, the difference was smallest for orthographically similar words (see Figure 7).

3.1.3 Strategies: Verbal working memory. There was no difference in reported talk-out-loud strategy between the group with more inner speech (10 out of 47) and the group with less inner speech (13 out of 46) ($\chi^2(1) = 0.29$, p = .589). 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 disappeared when they reported that they said the words out loud to help them

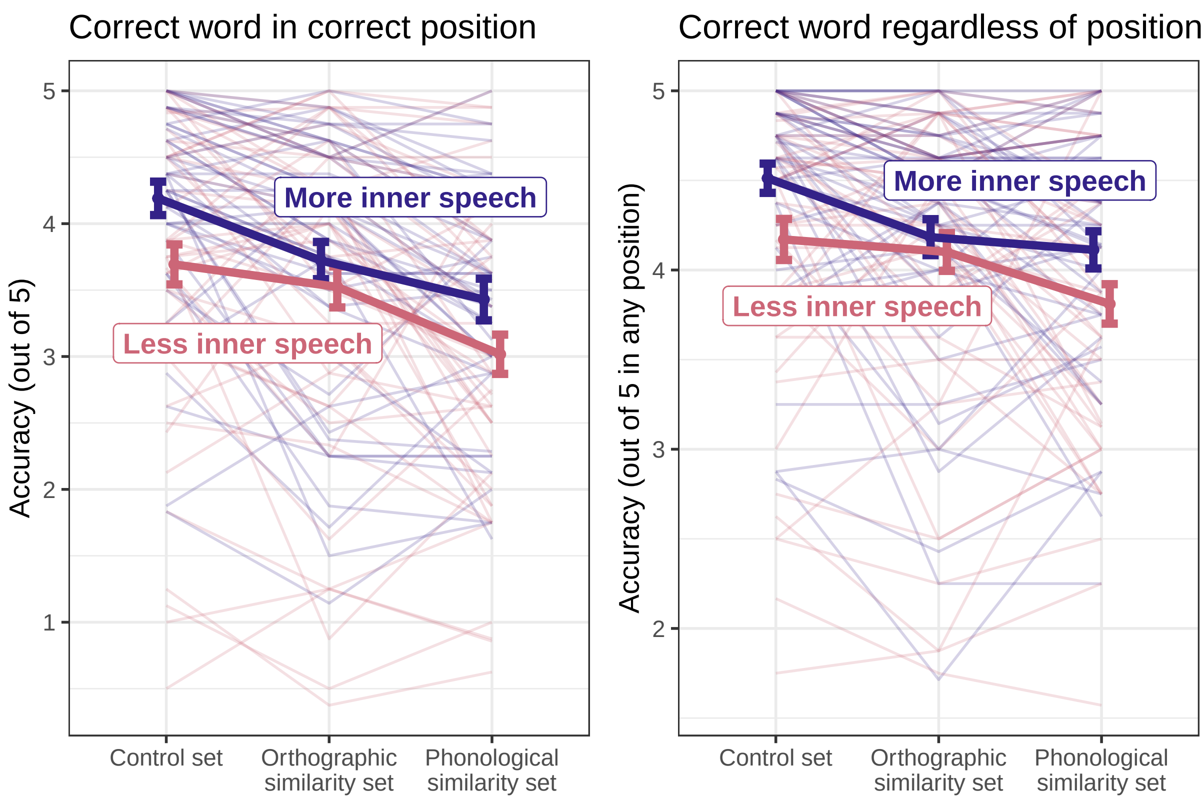


Figure 7. Score on the verbal working memory task by word set.

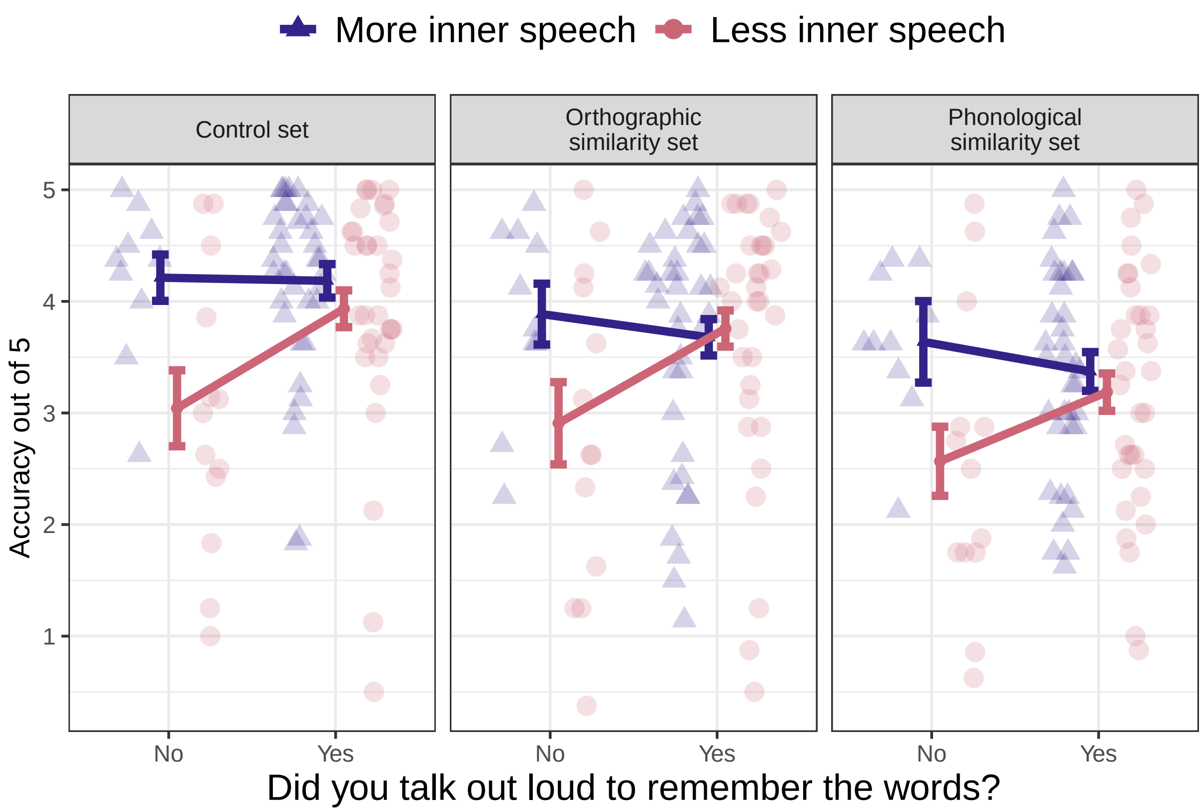


Figure 8. Verbal working memory performance by whether participants reported talking out loud to help them remember or not.

remember. Participants reporting more inner speech remembered the words better, but this effect was canceled out when participants reported talking out loud to solve the task (interaction effect: β = -0.50; SE = 0.23; t = -2.19; p = .031).

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. The below-chance performance was likely due to the low name agreement of at least one image in each pair (mean agreement rating for these 10 images = 0.58; range = 0.05 to 1).

3.2.1 Descriptive statistics by group: Rhyme judgments. As can be seen in Table 3, participants with more inner speech were generally both faster and more accurate than participants with less inner speech on all three types of trials. See also Figure 9.

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	82.77	2.86
More inner speech	No rhyme	1931	53	97.52	1.36
More inner speech	Orthographic rhyme	1719	55	91.21	2.48
Less inner speech	Non-orthographic rhyme	1970	54	76.20	3.21
Less inner speech	No rhyme	2024	60	93.84	1.87
Less inner speech	Orthographic rhyme	1859	60	83.62	3.22

3.2.2 Statistical models: Rhyme judgments. Participants took longer to make rhyme judgments on no-rhyme trials (M = 1981 ms) compared with orthographic

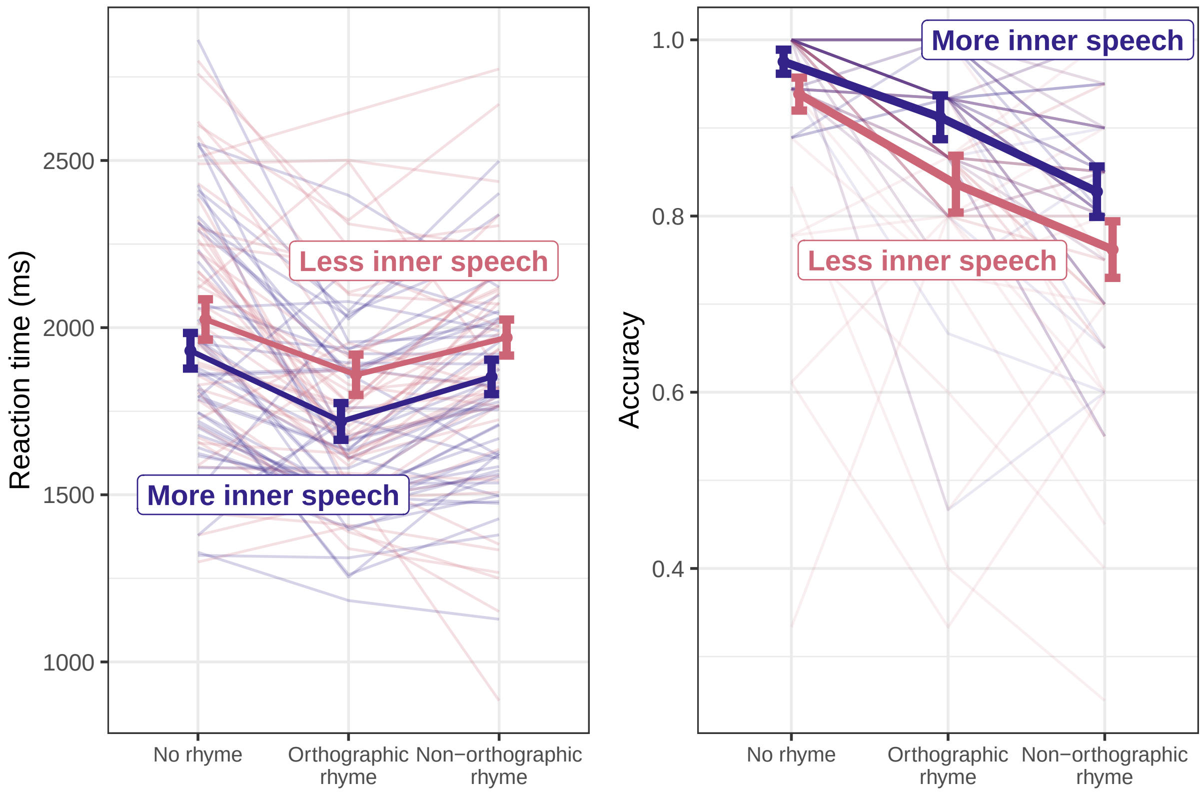


Figure 9. Reaction time and accuracy across groups by rhyme type.

trials (M = 1730 ms) (β = 0.12; SE = 0.04; t = 2.97; p = .005). This means that no-rhyme trials took 13% longer than orthographic trials ($e^{0.12} = 1.13$). Non-orthographic trials (M = 1821 ms) did not differ significantly from orthographic trials (β = 0.04; SE = 0.04; t = 1.11; p = .272). Trials where the presented images had higher name agreement were also faster (β = -0.04; SE = 0.02; t = -2.25; p = .029). Reported inner speech had no effect on speed of rhyme judgments (β = -0.02; SE = 0.02; t = -0.81; p = .422), and there were no interactions between rhyme type and verbal score (both $p > .298$). Verbal score and name agreement also did not interact ($p > .975$).

Participants were more accurate on no-rhyme judgments (M = 95.7%) than on orthographic rhyme judgments (M = 87.5%) (β = 1.30; SE = 0.29; z = 4.49; $p < .001$) and less accurate on non-orthographic rhyme judgments (M = 79.5%) than on orthographic rhyme judgments (β = -0.58; SE = 0.26; z = -2.18; p = .029). A higher verbal score was associated with a higher likelihood of responding accurately (β = 0.31; SE = 0.12; z = 2.57; p = .010). Trials with images with higher name agreement were not significantly easier ($p < .139$). There was no significant interaction between rhyme type and verbal score (both $p > .311$) or between verbal score and name agreement ($p = .324$).

3.2.3 Strategies: Rhyme judgments. There was no significant difference between how many participants with more inner speech (23 out of 47) and how many participants with less inner speech (21 out of 46) reported that they had said the words out loud ($\chi^2(1) = 0.01$, $p = .913$). Nevertheless, the effect of doing so was interestingly different for the two groups as can be seen in Figure 10. Saying the words out loud diminished the accuracy advantage associated with a higher verbal score for non-orthographic rhymes (β = -0.72; SE = 0.28; z = -2.53; p = .012) and orthographic rhymes (β = -0.69; SE = 0.31; z = -2.25; p = .024) compared with no-rhyme trials. This suggests that this was the strategy that participants with more inner speech used covertly.

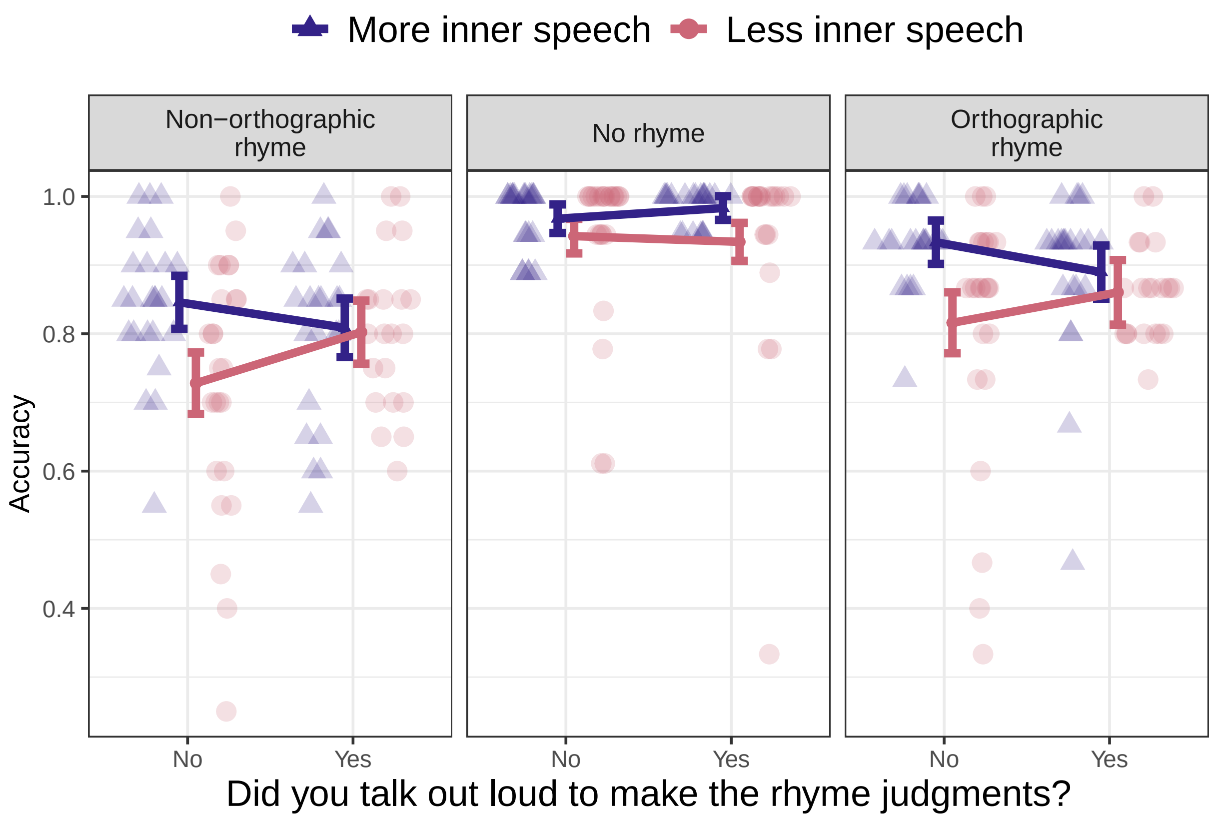


Figure 10. Reaction time and accuracy by whether participants indicated that they had talked out loud to make the rhyme judgments.

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 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.

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)
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

3.3.2 Statistical models: Task switching. Participants responded less accurately in the symbol-cued switch condition (M = 97.2%), in the color-cued switch

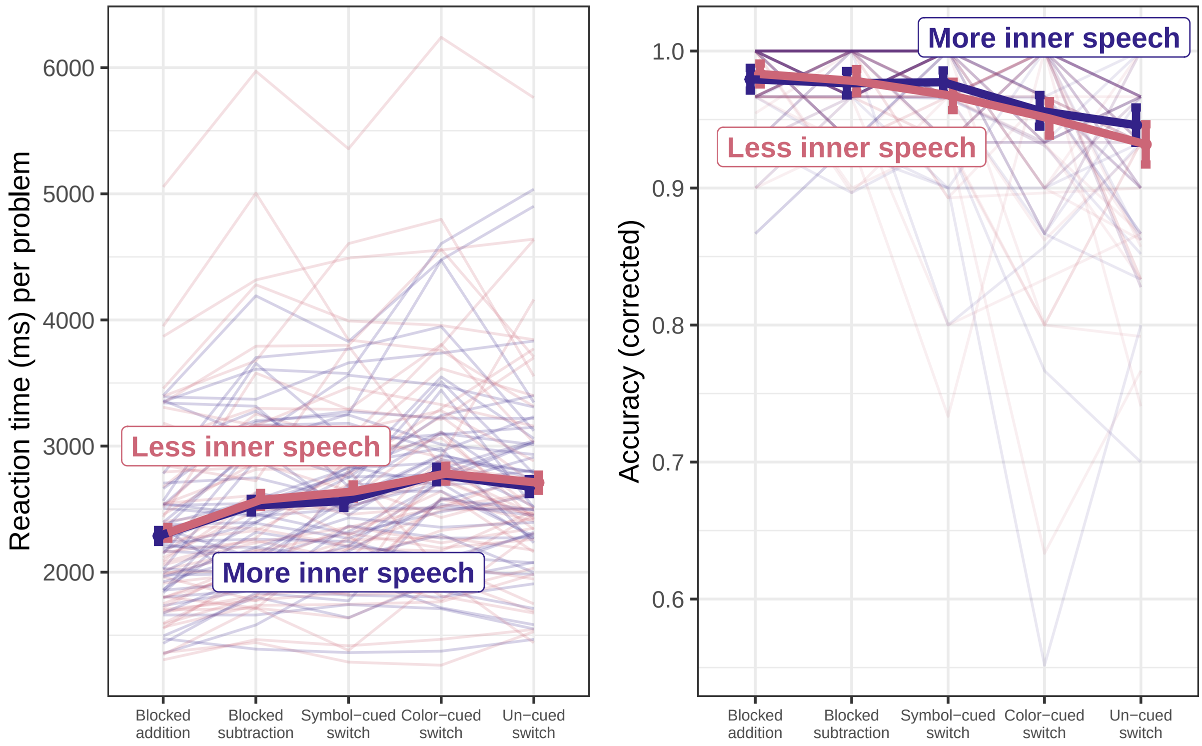


Figure 11. Reaction time and accuracy across conditions in the task switching experiment.

condition (M = 95.4%), and in the un-cued switch condition (M = 93.9%) compared with the blocked addition condition (M = 98.1%) (addition versus symbol-cue: β = -0.42; SE = 0.18; z = -2.32; p = .020; addition versus color-cue: β = -0.97; SE = 0.17; z = -5.84; p $<$ .001; addition versus un-cued: β = -1.27; SE = 0.16; z = -7.92; p $<$ .001). Accuracy did not differ between blocked subtraction (M = 97.7%) and blocked addition (p = .239). More inner speech was not associated with different accuracy (p = .547) and there were no interaction effects between inner speech and block-type (all p $>$ .075). 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 with the subtraction condition (M = 2550 ms) (β = 0.09; SE = 0.01; t = 8.41; p $<$ .001; regression coefficient: $e^{0.09} = 1.09$), the symbol-cued switch condition (M = 2601 ms) β = 0.12; SE = 0.01; t = 9.69; p $<$ .001; regression coefficient: $e^{0.12} = 1.13$), the color-cued switch condition (M = 2778 ms) (β = 0.19; SE = 0.02; t = 12.23; p $<$ .001; regression coefficient: $e^{0.19} = 1.21$), and the un-cued switch condition (M = 2694 ms) (β

= 0.15; SE = 0.02; t = 9.39; p < .001; regression coefficient: $e^{0.15}$ = 1.16). More reported inner speech did not predict reaction times (p = .810), and there were no interaction effects (all p > .516).

3.3.3 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 = .318). 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).

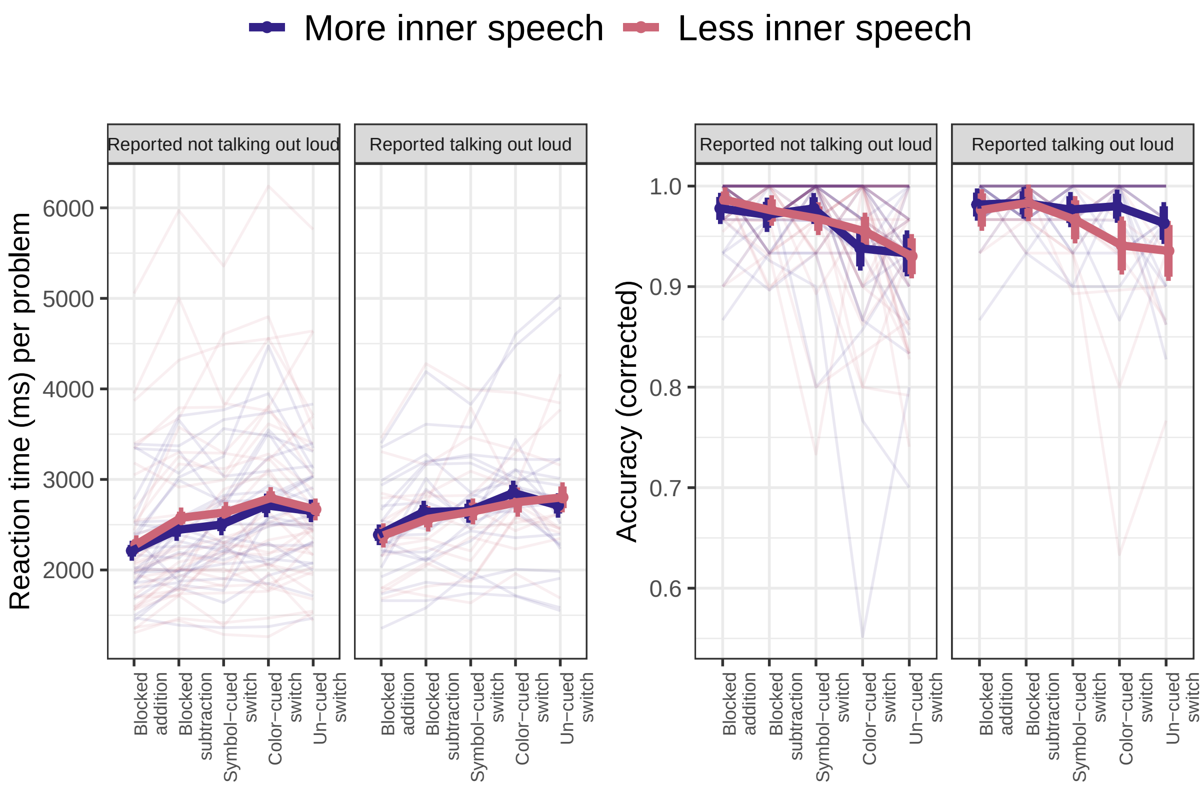


Figure 12. Reaction time (ms) and accuracy in the task switching experiment by whether participants reported talking out loud to remember the correct rule or not.

3.4 Same/different judgments

We excluded trials above 5 seconds (0.7 %) and below 200 ms (0.07 %). Generally, participants made the correct judgment on 95.53 % of trials. This did not differ between the group of participants with more inner speech (95.58 %) and the group with less inner speech (95.48 %). In subsequent analyses and plots, we only include correct trials.

3.4.1 Descriptive statistics by group: Same/different judgments.

See Figure 13 for reaction times between the groups with more inner speech and less inner speech for category judgments (‘do these two animals belong to the same category?’) or identity judgments (‘are these two animals identical?’).

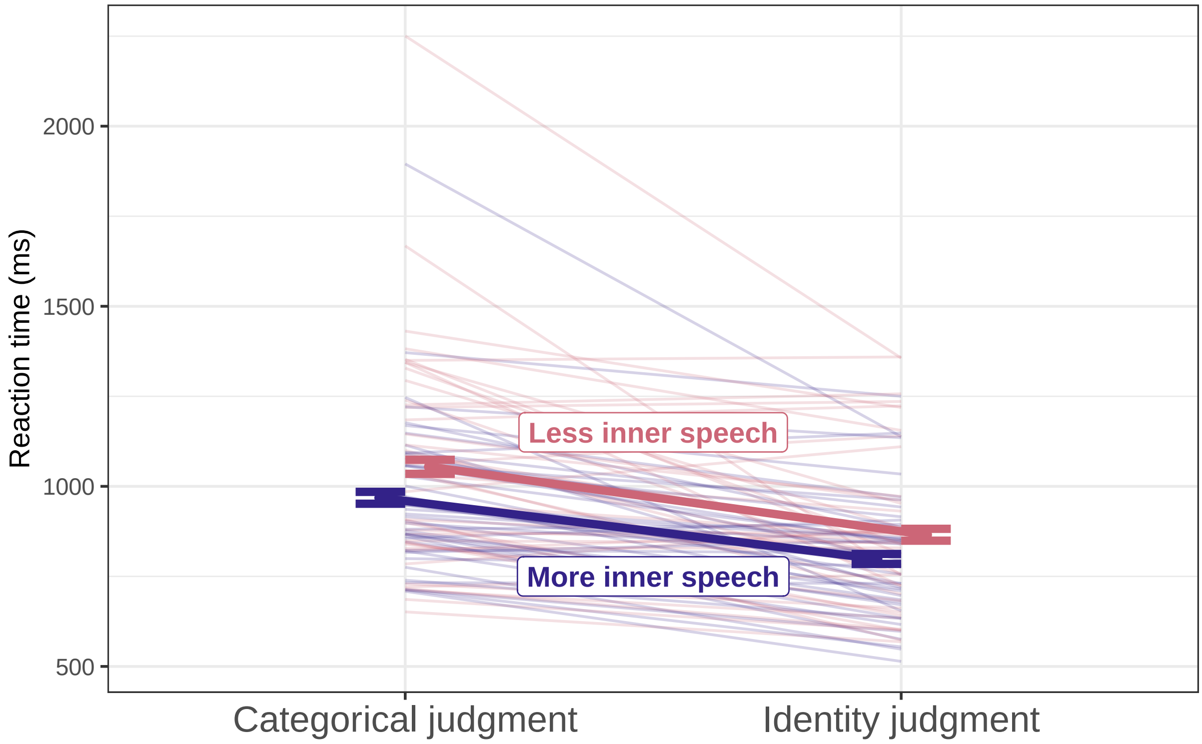


Figure 13. Reaction time in response to category or identity judgments.

3.4.2 Statistical models: Same/different judgments.

Identity judgments (M = 832 ms) were faster than category judgments (M = 1010 ms) (β = -0.19; SE = 0.02; t = -11.38; p < .001; regression coefficient: $e^{-0.19} = 0.83$), and a higher verbal score was not associated with faster reaction times (β = -0.03; SE = 0.02; t = -1.57; p = .120; regression coefficient: $e^{-0.03} = 0.97$).

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). See Figure 14. However, participants with more inner speech were not specifically adversely affected by the

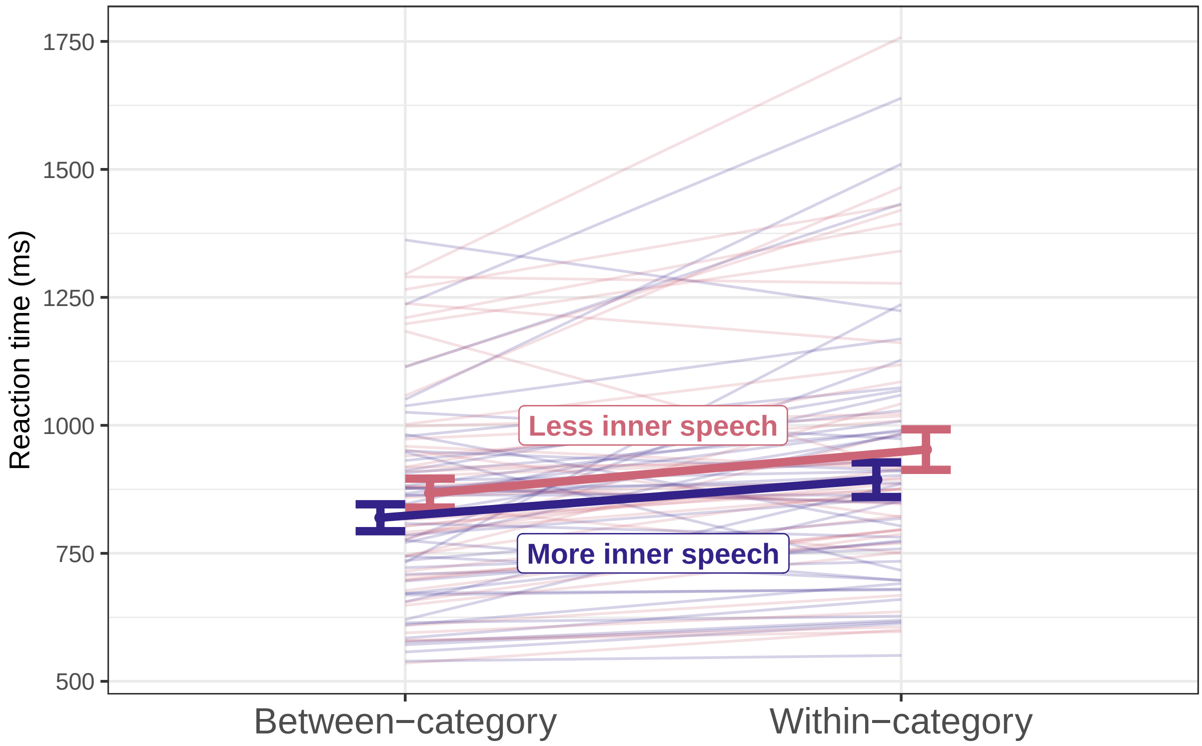


Figure 14. Reaction time on identity trials where the correct response was 'DIFFERENT' either because the two silhouettes were from different categories or different images from the same category.

within-category interference (interaction effect: (β = 0.00; SE = 0.01; t = -0.06; p = .954). 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; regression coefficient: $e^{-0.08}$ = 0.92). ### 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 switching experiment ($\chi^2(1)$ = 1.33, p = .248). There were not any differences between the effects that talking out loud had on these two groups.

3.5 Questionnaire measures

Because of a technical error, we are missing questionnaire data from one participant from the group with less inner speech, so we here report questionnaire data from 47 participants with more inner speech and 45 participants with less inner speech. For most of our custom questions, there were notable differences in how participants from the two groups responded. For reasons of space, however, we only report a few illustrative ones here (see Appendix D for plots of all the questions). 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 (see Figure 15) (revise past conversation: $t(87.95) = 5.93$; $p < .001$; practice future conversation: $t(89.33) = 5.33$; $p < .001$). Of the VISQ factors, our verbal representation score was mostly related to the dialogicality of inner speech (see again Figure 15) ($r(90) = .70$; $p < .001$).

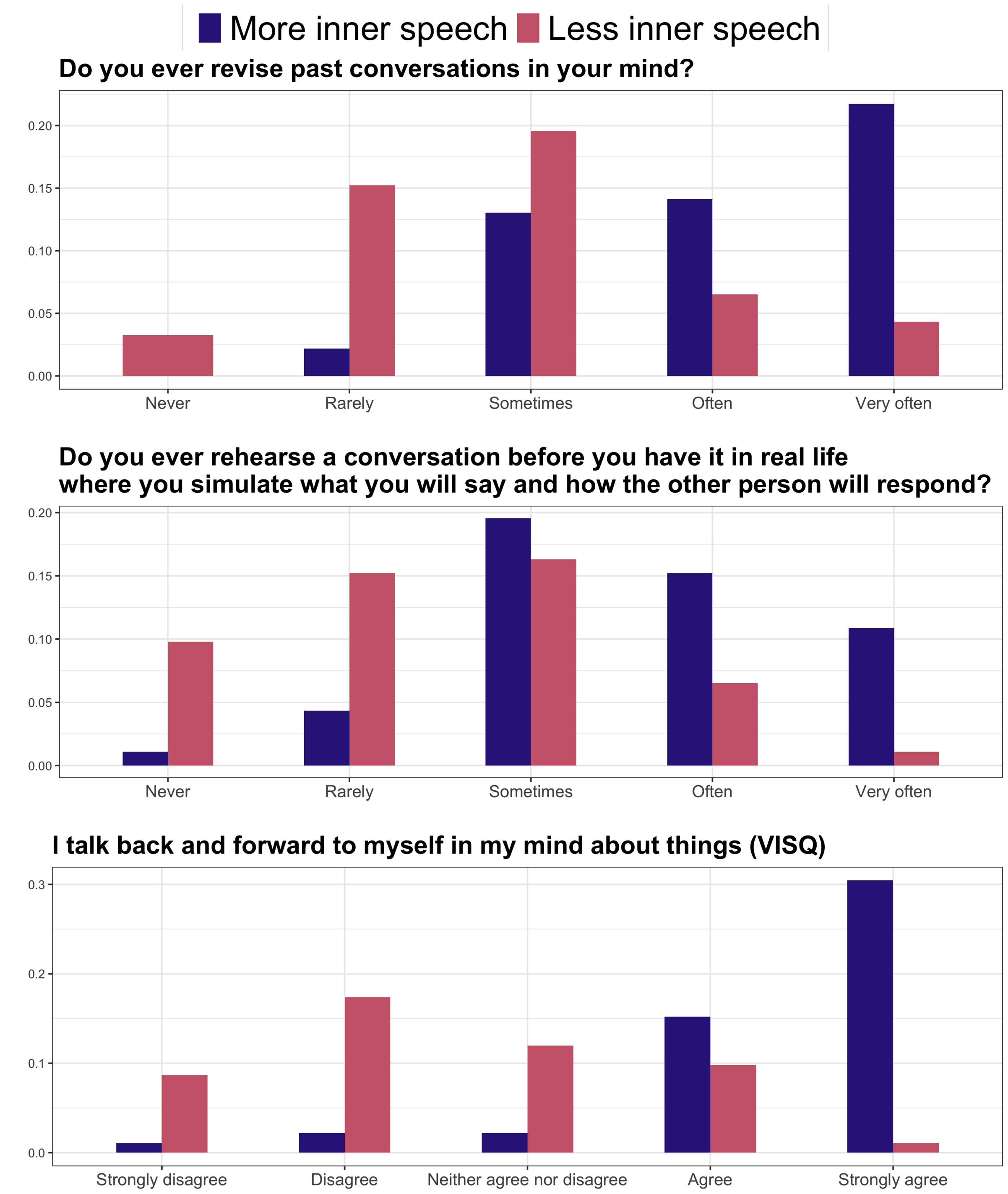


Figure 15. Grouped bar plots of proportional answers to selected custom questions concerning inner speech. Dark blue represents participants with more inner speech, and pink represents participants with less inner speech.

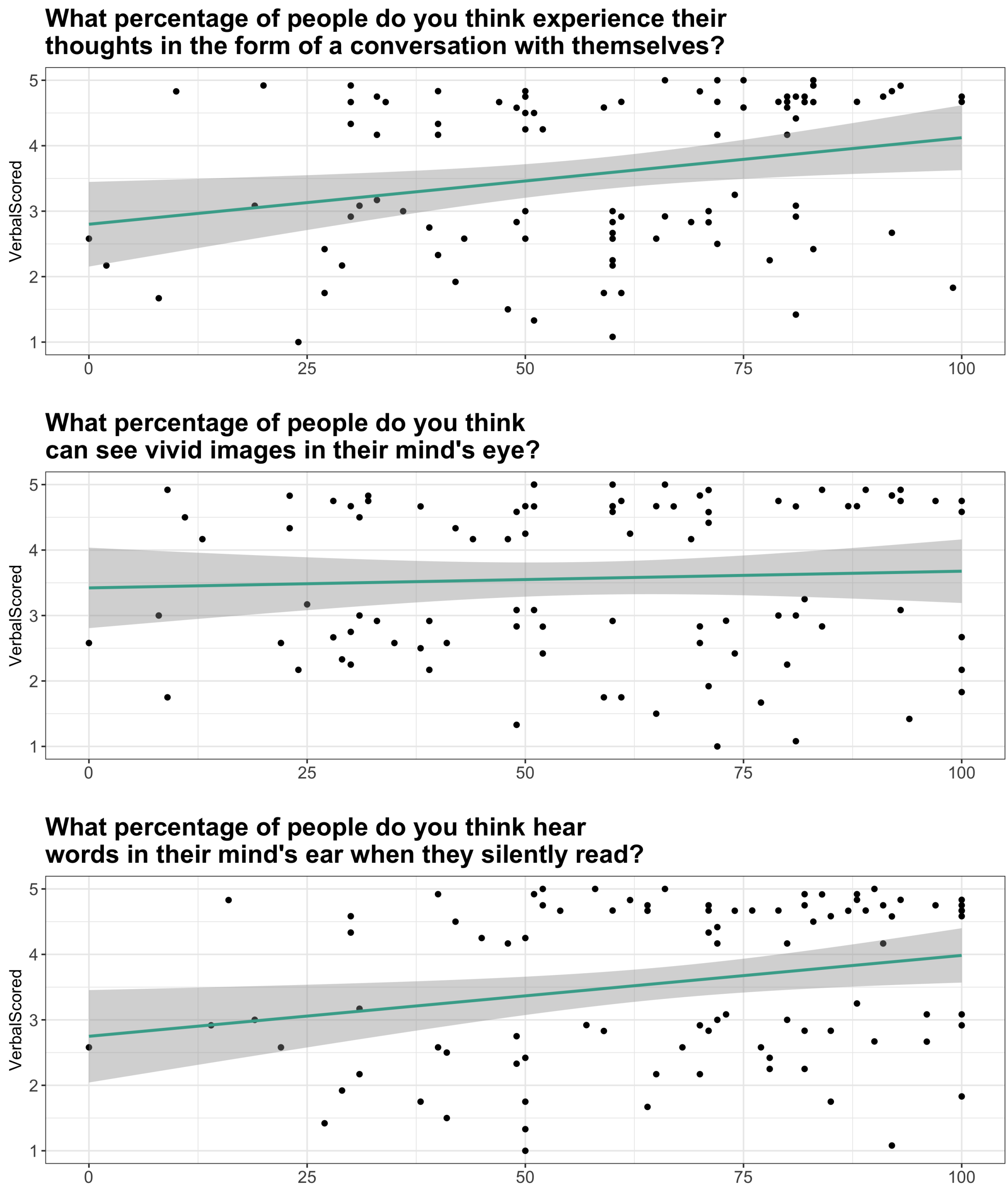


Figure 16. Scatter plots showing the correlation between verbal score on the IRQ and participants' estimates of percentages of other people with a given kind of experience.

It was also remarkable that participants' own experience influenced how they thought other people's experience was (see Figure 16). Participants who reported more inner speech estimated that more people generally experience their thoughts in the form of a conversation with themselves ($\beta = 5.08$; SE = 2; t = 2.55; p = .013) and that more people generally hear words in their "mind's ear" when they read ($\beta = 5.09$; SE = 2.07; t = 2.46; p = .016). 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 = .605).

4 Discussion

Participants who report experiencing less inner speech (our sample targeted those at $<$ 16%ile of the verbal score on the IRQ) differed in performance on several behavioral measures. First, they displayed poorer verbal working memory regardless of the material. However, contrary to our prediction, there was no indication of a weaker (or stronger) phonological similarity effect as a function of inner speech. Second, participants who report less inner speech were less accurate at judging whether the names of two images rhymed. The lack of an inner speech by nameability interaction makes it more likely that the effect stemmed from comparing phonological representations in memory rather than naming the images themselves. 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, suggesting a kind of compensatory mechanism. Inner speech differences did not predict performance in task switching which is somewhat surprising given substantial previous research showing endogenously cued task switching being susceptible to verbal interference (Nedergaard et al., 2022). 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 makes the use of inner speech unlikely. In terms of our custom questionnaire, participants responded in ways consistent with their IRQ answers. Participants with more inner speech were for example more likely to rehearse past and future conversations and to estimate that others experience an inner voice when they read and that other people experience their thoughts in the form of a conversation with themselves.

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, the term *endophasia* already exists as a term for inner speech (Bergounioux, 2001; Loevenbruck et al., 2018). The term also avoids subsuming a lack of inner speech under “aphantasia” (Monzel, Mitchell, Macpherson, Pearson, & Zeman, 2022) which we would like to avoid 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 (Bermúdez, 2018; Gauker, 2018; Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Loevenbruck, 2014). 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?

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 their experience rarely takes a verbal format, they are not just confabulating. This is evident both in the consistency of their subjective responses (Roebuck & Lupyan, 2020), and, as we report here, there are some clear behavioral correlates. 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 substrates (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016). This 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 does not appear to be (Dance, Ipser, & Simner, 2022).

We did find evidence that using other strategies than internal verbalization could reduce 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 could be the reason why there was reduced difference in performance between the two groups for this word set. Similarly, the finding that the two groups did not differ in either reaction time or accuracy on the task switching experiment could suggest that while the inner voice can be used as a behavioral self-cue, other and equally effective strategies may be available. As mentioned in the Introduction, different strategies resulting in similar behavioural outcomes have also been found studies of people with aphantasia (Keogh et al., 2021).

4.2 Relations to visual imagery, condensed inner speech, and unsymbolized thought

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. Our participants with anendophasia were not totally unable to make covert rhyme judgments, for example, as even the participants who reported not naming the images out loud performed above chance. They just found the task more difficult than the participants with more inner speech did. Instead of a total inability to verbalize internally, 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). Given that individuals with anendophasia had issues specifically with tasks that required storage and comparisons of phonological representations, it could be the case that they experience a kind of inner speech without "hearing" the speech sounds or "feeling" the articulation. Indeed, some individuals from the online communities reported that they do experience words but not the sounds of words when they think. Of the participants in the present study, one 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'. The informal reports of individuals with anendophasia thus parallel findings from Descriptive Experience Sampling of both "wordless" inner speech and unsymbolized thinking, akin to "thinking in ideas". These kinds of inner speech can potentially be usefully conceptualised as different levels of condensation of inner speech (Vicente & Martinez-Manrique, 2016). It still remains an open question whether individuals with anendophasia experience highly condensed inner speech with attenuated imagery or purely unsymbolized thought. This question could potentially be addressed through studies investigating whether or how much the categories of natural language influence individuals with anendophasia. Such study designs could for example be inspired by color categorization studies (Gilbert, Regier, Kay, & Ivry, 2006, 2008; Winawer et al., 2007).

4.3 Limitations of the present study

One limitation of our work is its reliance on wholly subjective questions for measuring inner speech. Considering that our focus is on the behavioral correlates of 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 are often wrong when they report their experience (Hurlburt & Schwitzgebel, 2011), especially if such reports take place retrospectively and require interpretations (Berger, Dennehy, Bargh, & Morsella, 2016; Ericsson & Simon, 1980; Nisbett & Wilson, 1977). It would therefore be helpful to supplement subjective assessments with physiological measures of the sort becoming possible for differences in visual imagery like investigating priming with binocular rivalry (Keogh & Pearson, 2018) or effects of visual imagery on pupil dilation (Kay, Keogh, Andrillon, & Pearson, 2022). Another limitation is the remaining possibility that differences we ascribe to inner speech come from something else such as differences in conscientiousness. We believe this is unlikely since we saw examples of specific conditions where there were no differences between the two groups (e.g., no-rhyme pairs, orthographically similar words, and all conditions in the task switching experiment). However, future studies could include separate measures of conscientiousness (e.g., using the Big Five Inventory, John, Donahue, & Kentle, 1991) and general intelligence, insofar as such exists (e.g., using Raven's progressive matrices, Raven, 2000).

4.4 Future directions

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 (see e.g., Keogh et al., 2021). We see some indications in our present study as well with the different effects of using a talk-out-loud strategy for the two groups. Such strategy differences should be explored in future studies, ideally through experiments where different strategies would show different behavioral profiles.

If it is correct that what people with anendophasia experience is highly condensed inner speech rather than no inner speech at all, this would also lead to predictions about the functions of their inner speech. For example, they should be less likely to use it in contexts where the specific words and sounds are important, such as as a mnemonic aid (e.g., for rehearsing a shopping list) or for simulating conversations. Indeed, the most striking difference between the two groups in the questionnaire was that participants with more inner speech spent more time rehearsing past and future conversations which makes us wonder what kind of consequences this might have. Would we expect people with more inner speech to be somehow "better" at conversations? Or maybe worse because they over-rehearse? It seems that inner speech is 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) where participants judge situations with social mishaps (did a faux pas occur and, if so, why was it a faux pas). This would indicate whether differences in inner speech use are related to social abilities not strictly reliant on communication.

5 Conclusion

Not everyone experiences inner speech. We proposed a name for a lack of the experience of inner speech: anendophasia. Participants with anendophasia were worse at making rhyme judgments in response to images and remembering a list of words. However, they did not differ from the control group in either task switching performance or visual discrimination judgments. They reported less auditory imagery generally (e.g., had songs stuck in their heads less often) and otherwise responded to our custom questionnaire in ways consistent with less propensity to engage in habitual inner speech. 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 due to people with anendophasia using alternative strategies. It is an open question whether anendophasia is actually a lack of inner speech or simply a lack of the experience of inner speech because of weak or absent articulatory-auditory imagery.

6 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.

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

Berger, C. C., Dennehy, T. C., Bargh, J. A., & Morsella, E. (2016). Nisbett and wilson (1977) revisited: The little that we can know and can tell. *Social Cognition*, *34*(3), 167–195.

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.

Bermúdez, J. L. (2018). Inner Speech, Determinacy, and Thinking Consciously about Thoughts. In *Inner Speech: New Voices* (Vol. 1, pp. 199–220). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0008

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

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

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.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable: Export tables to LaTeX or HTML*. Retrieved from https://CRAN.R-project.org/package=xtable

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.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

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

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.

Gauker, C. (2018). Inner Speech as the Internalization of Outer Speech. In *Inner Speech: New Voices* (Vol. 1, pp. 53–77). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0003

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.

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Attention & Performance XVIII: Control of Cognitive Processes*. Cambridge, MA: MIT Press.

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

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

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

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., 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, 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.

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

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

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

Kassambara, A. (2021). *Rstatix: Pipe-friendly framework for basic statistical tests*. Retrieved from https://CRAN.R-project.org/package=rstatix

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., & 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., 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.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

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

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

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

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

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

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.

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/

Raven, J. (2000). The raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48.

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.

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

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

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

Vygotsky, L. S. (1962). *Thought and language*. Cambridge: Massachusetts Institute of Technology. New York; London: John Wiley.

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., Henry, L., Pedersen, T. L., Luciani, T. J., Decorde, M., & Lise, V. (2021). *Svglite: An 'SVG' graphics device*. Retrieved from https://CRAN.R-project.org/package=svglite

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.

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

7 Supplemental materials

7.1 Appendix A: Materials for the rhyme judgment experiment

See 'rhyming_images' folder for the images used. See Table 5 for all image files used along with their name agreement scores from a separate experiment.

Table 5

Image files for the rhyme judgment experiment including name agreement (0 to 1) from a separate validation experiment.

File name	Name agreement
bag.png	0.95
bear.png	0.95
bed.png	0.95
beer.png	0.75
bell.png	0.90
bin.png	0.10
bone.png	0.95
boot.png	0.95
box.png	0.95
brain.png	0.95
bread.png	0.90
cab.png	0.05
cake.png	0.05
cat.png	0.90
cave.png	0.55
chain.png	0.95
chair.png	0.95
chess.png	0.25
chin.png	0.80
claw.png	0.55
clock.png	0.95
cone.png	0.70
crab.png	0.95
crane.png	0.70
dart.png	0.90
deer.png	0.65
dog.png	0.95
door.png	0.95
drawer.png	0.80
dress.png	0.85

Table 5

Image files for the rhyme judgment experiment including name agreement (0 to 1) from a separate validation experiment. (continued)

File name	**Name agreement**
drum.png	1.00
egg.png	0.95
eye.png	0.95
fan.png	0.75
flag.png	0.90
fly.png	0.95
fox.png	0.85
hair.png	0.90
hat.png	1.00
heart.png	0.90
house.png	0.95
jar.png	0.95
key.png	0.95
king.png	0.80
lab.png	0.70
leg.png	0.85
man.png	0.80
moon.png	0.95
mouse.png	0.85
nail.png	0.95
nose.png	1.00
park.png	0.70
pear.png	0.95
plane.png	0.35
pope.png	0.45
rake.png	1.00
ring.png	0.90
rope.png	1.00
rose.png	0.95
saw.png	1.00
screw.png	1.00
seal.png	0.75
shark.png	0.75
shell.png	0.15
shoe.png	0.85
snail.png	0.95

Table 5

Image files for the rhyme judgment experiment including name agreement (0 to 1) from a separate validation experiment. (continued)

File name	Name agreement
soap.png	0.80
sock.png	0.90
socks.png	0.90
spoon.png	1.00
square_rhyme.png	0.85
star.png	0.95
suit.png	0.95
thumb.png	0.90
tie.png	0.70
train.png	0.95
tree.png	0.85
triangle.png	1.00
wave.png	0.85
well.png	0.95
whale.png	0.95
wheel.png	0.75

7.2 Appendix B: Custom questionnaire items

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)
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)

(continued)

Question	Options
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)
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)

(continued)

Question	Options
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)
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)

(continued)

Question	Options
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%)

7.3 Appendix C: R packages

R packages used: R (Version 4.3.0; R Core Team, 2022) and the R-packages *corrplot2021* (Wei & Simko, 2021), *cowplot* (Version 1.1.1; Wilke, 2020), *data.table* (Version 1.14.8; Dowle & Srinivasan, 2021), *dplyr* (Version 1.1.2; Wickham, François, Henry, & Müller, 2021), *forcats* (Version 1.0.0; Wickham, 2021a), *Formula* (Version 1.2.5; Zeileis & Croissant, 2010), *ggforce* (Version 0.4.1; Pedersen, 2021), *ggplot2* (Version 3.4.2; Wickham, 2016), *ggpubr* (Version 0.6.0; Kassambara, 2020), *Hmisc* (Version 5.0.1; Harrell Jr, Charles Dupont, & others., 2021), *kableExtra* (Version 1.3.4; Zhu, 2021), *lattice* (Version 0.21.8; Sarkar, 2008), *lme4* (Version 1.1.32; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *lubridate* (Version 1.9.2; Grolemund & Wickham, 2011), *Matrix* (Version 1.5.4; Bates & Maechler, 2021), *optimx* (Nash, 2014; Version 2022.4.30; Nash & Varadhan, 2011), *papaja* (Version 0.1.1; Aust & Barth, 2022), *purrr* (Version 1.0.1; Henry & Wickham, 2020), *readr* (Version 2.1.4; Wickham, Hester, & Bryan, 2021), *rstatix* (Version 0.7.2; Kassambara, 2021), *stringr* (Version 1.5.0; Wickham, 2019), *survival* (Version 3.5.5; Terry M. Therneau & Patricia M. Grambsch, 2000), *svglite* (Version 2.1.1; Wickham, Henry, et al., 2021), *tibble* (Version 3.2.1; Müller & Wickham, 2021), *tidyr* (Version 1.3.0; Wickham, 2021b), *tidyverse* (Version 2.0.0; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *trackdown* (Kothe, Callegher, Gambarota, Linkersdörfer, & Ling, 2021), *tufte* (Version 0.12; Xie & Allaire, 2022), and *xtable* (Version 1.8.4; Dahl, Scott, Roosen, Magnusson, & Swinton, 2019).

7.4 Appendix D: Custom questionnaire results

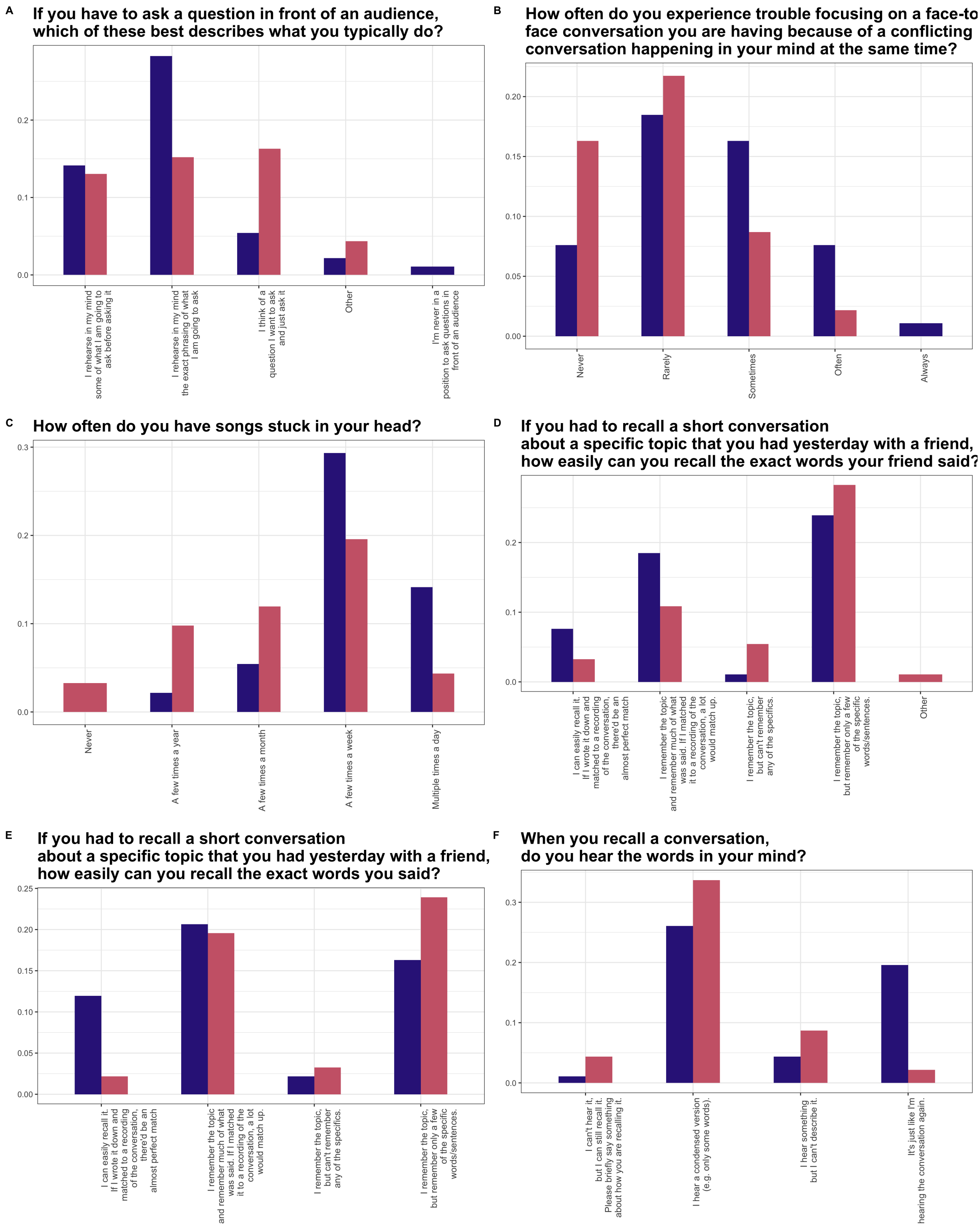


Figure 17. Grouped bar plots showing proportional answers (dark blue = more inner speech group; pink = less inner speech group).

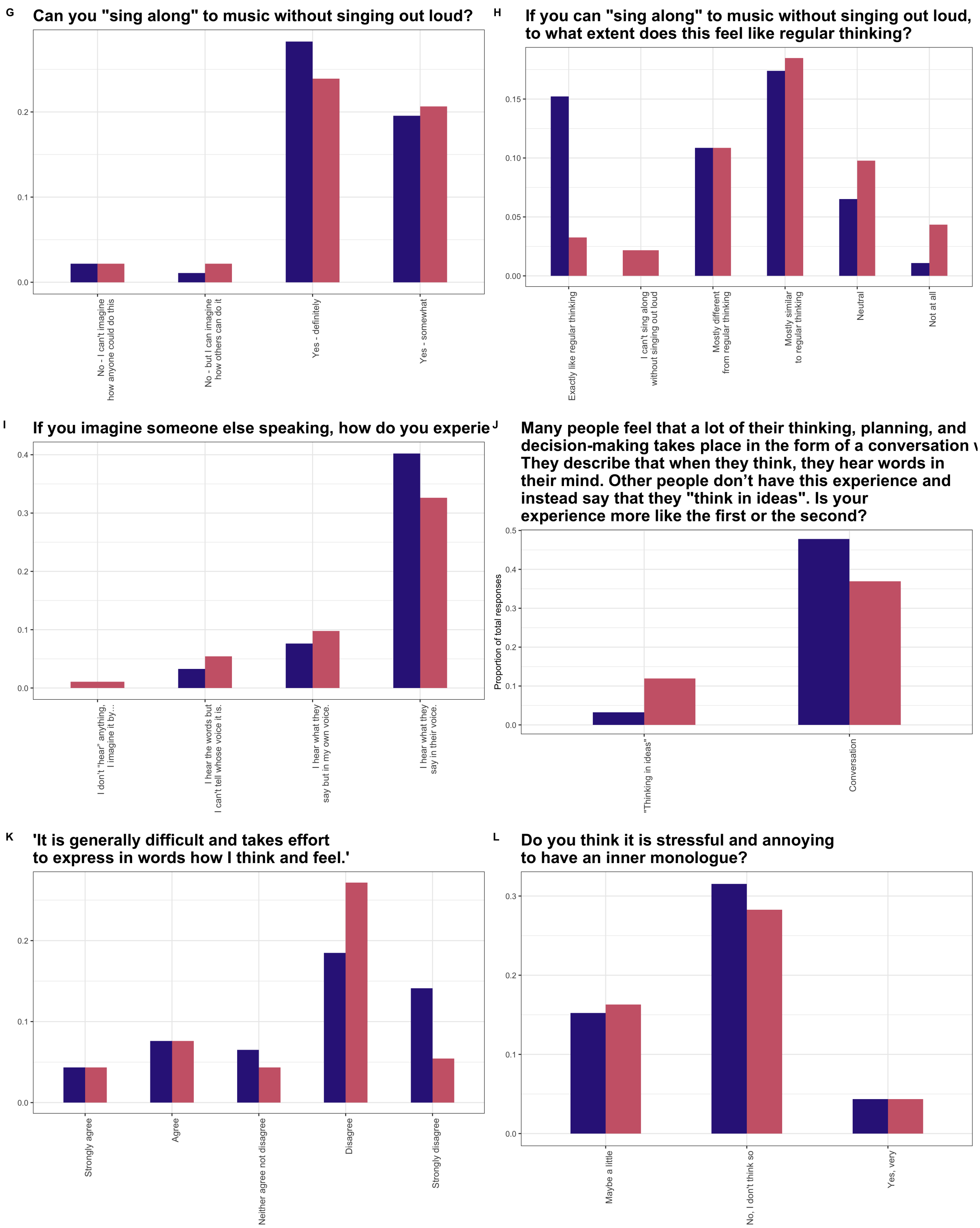


Figure 18. Grouped bar plots showing proportional answers (dark blue = more inner speech group; pink = less inner speech group).

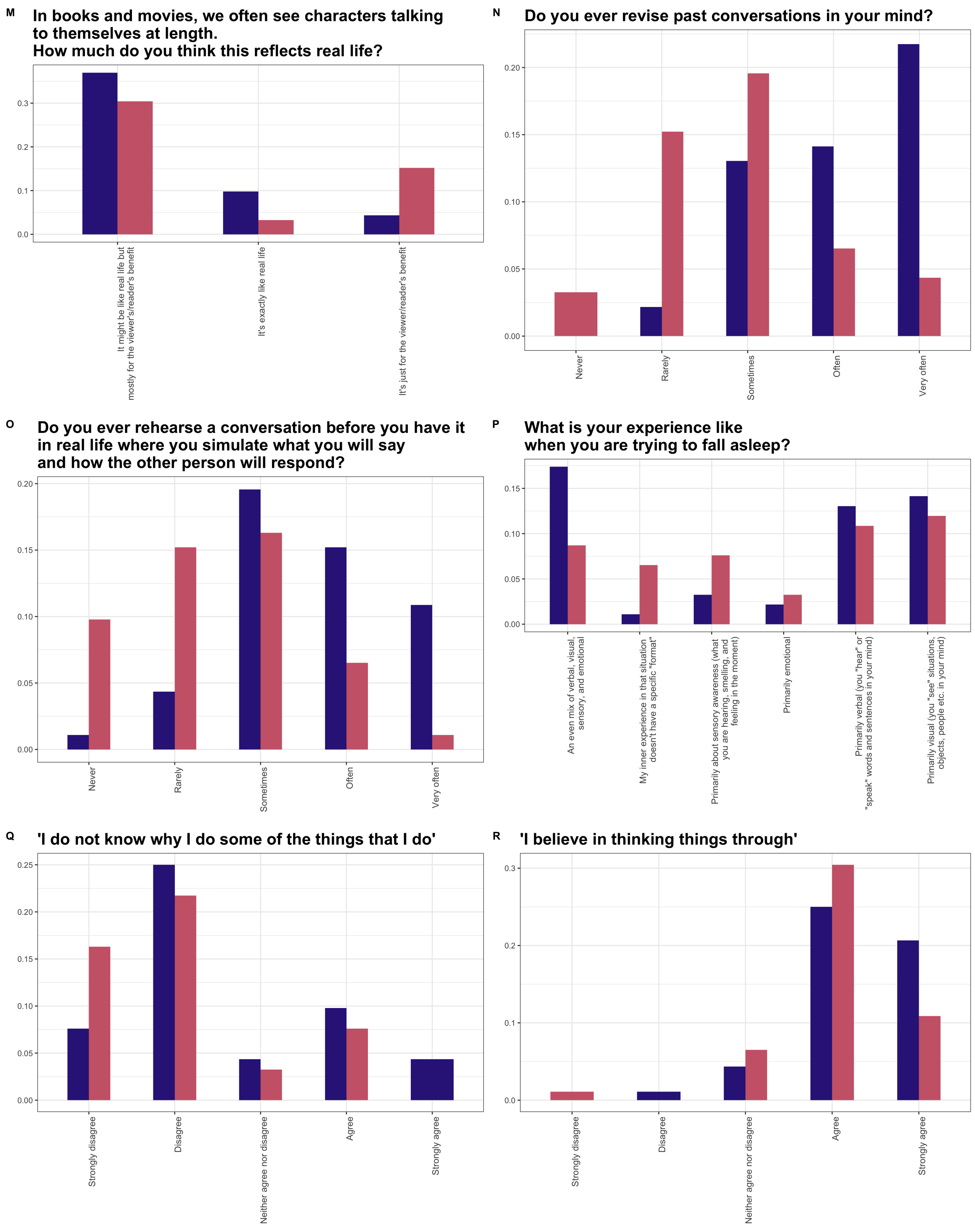


Figure 19. Grouped bar plots showing proportional answers (dark blue = more inner speech group; pink = less inner speech group).

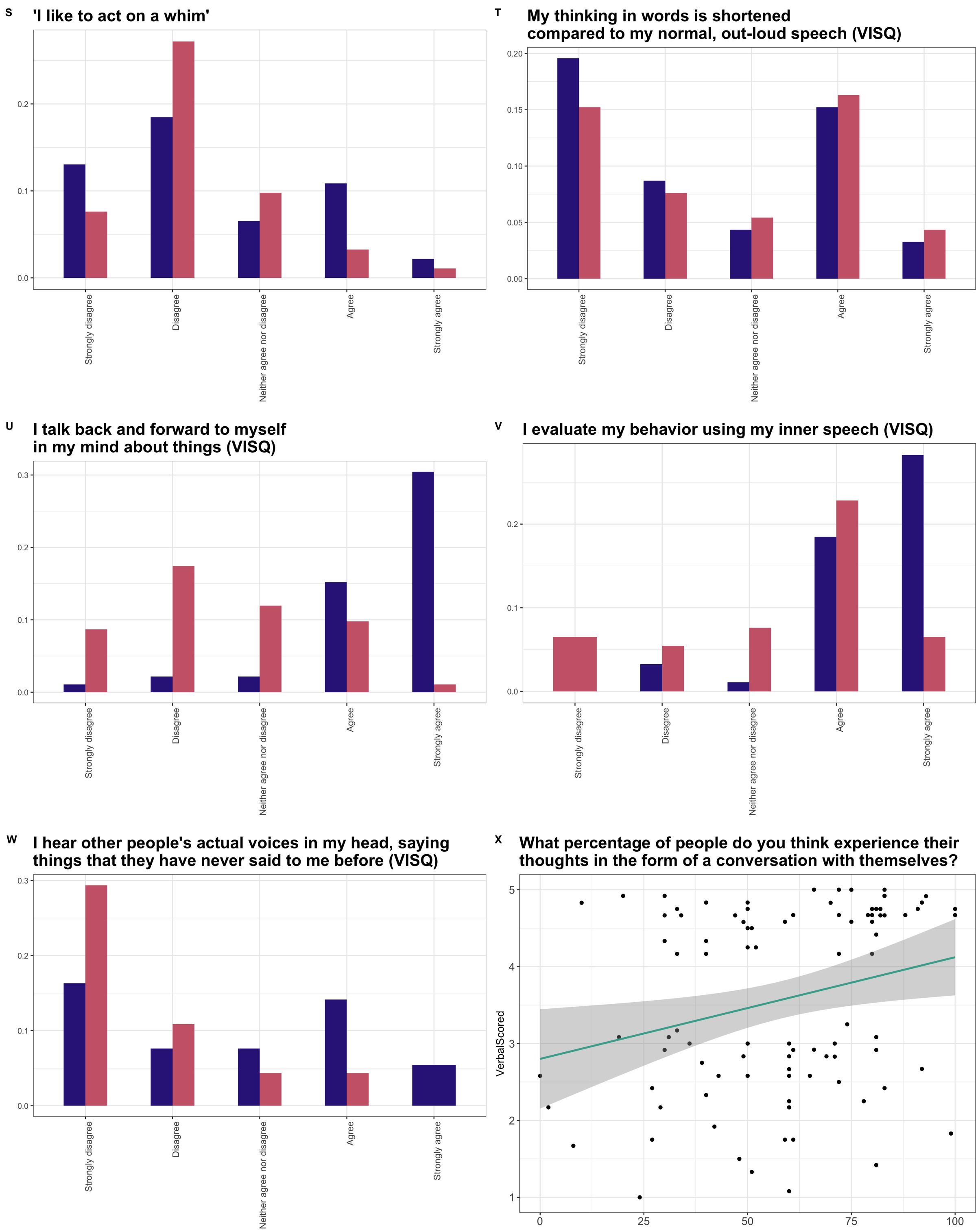


Figure 20. Grouped bar plots showing proportional answers (dark blue = more inner speech group; pink = less inner speech group).

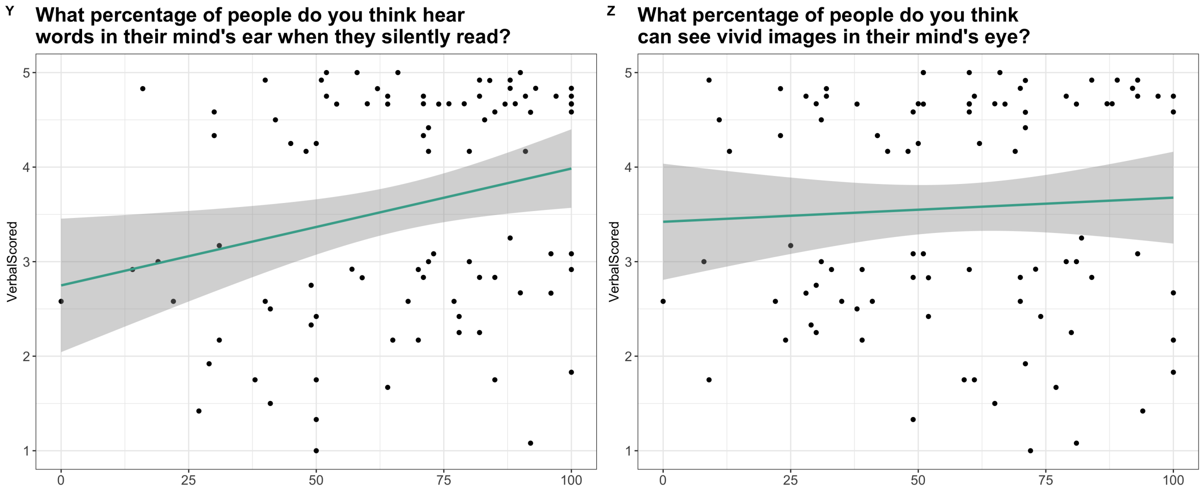


Figure 21. Figure S-W: Grouped bar plots showing proportional answers (dark blue = more inner speech group; pink = less inner speech group). Figures X-Z: Scatter plots showing correlation between verbal score on the IRQ and participants' estimates of percentages of other people with a given kind of experience.