‘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, Denmark

Word count: 7886

Author note

Johanne Nedergaard https://orcid.org/0000-0001-6381-2901

Mikkel Wallentin https://orcid.org/0000-0002-5509-9621

All experiment code and data can be accessed online at the Open Science Forum via this link:

https://osf.io/jgx7m/?view_only=10e0eb311dfc4ddd87e4a393ed2235e0

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

Abstract

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 minute 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.

Keywords

sustained attention; behavioral control; inner speech; mind-wandering

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.

Introduction

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 et al., 2022), Nedergaard and Wallentin (2021) 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 of better cycling performance) from being able to talk to themselves (see Nedergaard et al., 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, et al., 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, Morin, Faulds, Hall, and Wilson (2012) for example found that inner speech occurred in 60 percent 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 2) 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 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 towards 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

Openness and transparency 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/?view_only=10e0eb311dfc4ddd87e4a393ed2235e0. 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 modelled 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 % (95 % CI = [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. 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 round combined for brevity and clarity. See supplemental materials for analyses separated into first and second round. See Table 1 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.

Table 1. *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 to 83)
Median time spent (range)	19 mins 5 secs (15 mins 22 secs to 62 mins 25 secs)

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 (grey background).

Procedure

Participants first saw an instruction screen which 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 inter-stimulus 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 inter-stimulus intervals was randomized for each participant (30 seconds, 50 seconds, 60 seconds, 70 seconds, 120 seconds, 120 seconds, 150 seconds, 180 seconds). If they failed to respond to a circle prompt within 1 second, 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 2 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.

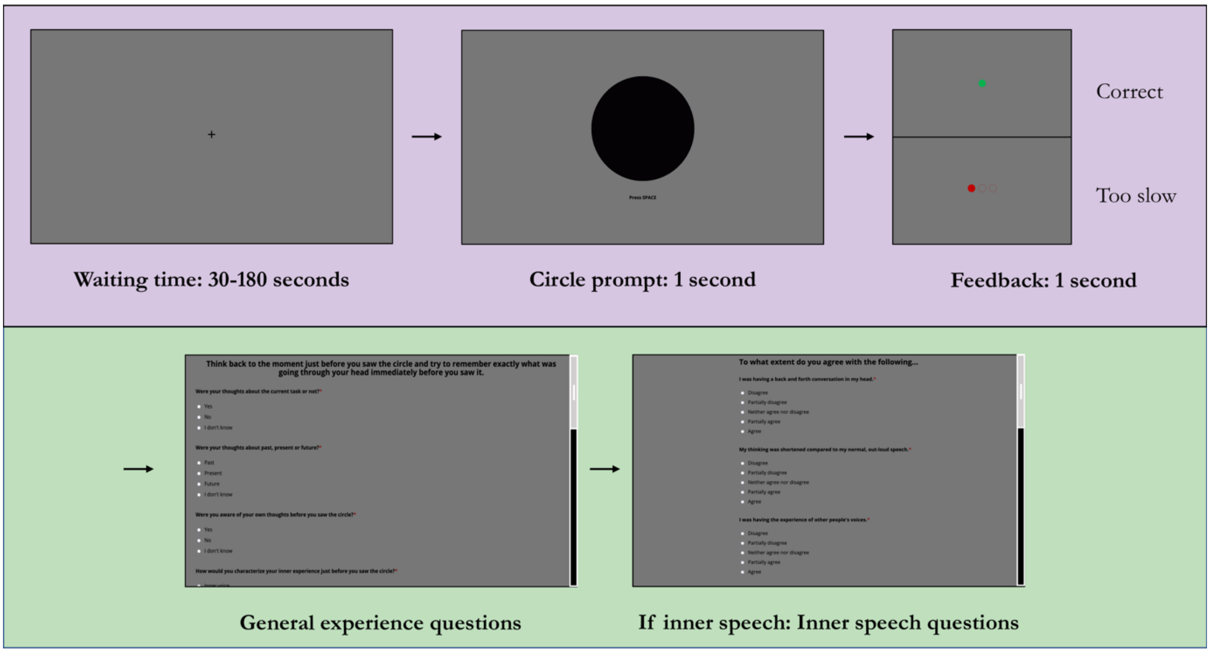

Figure 1. *Illustration of the experiment progression.*

Table 2. *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.*

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.

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 modelling, 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 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 < 1000 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).

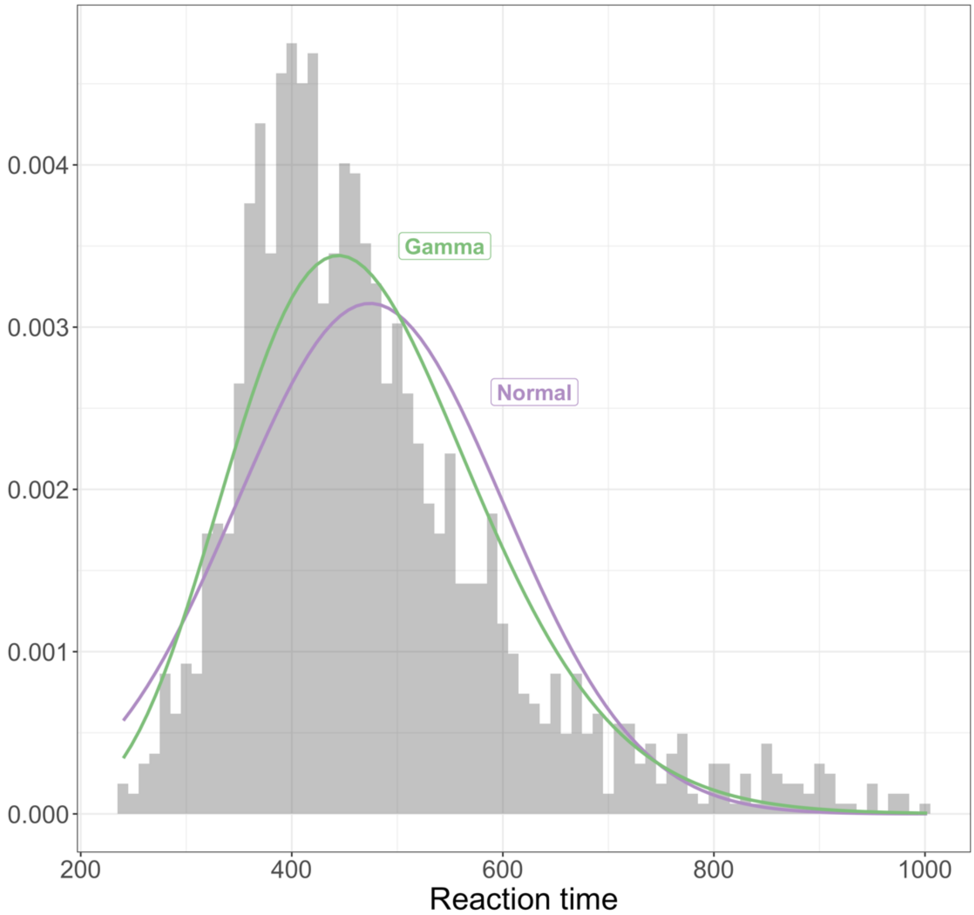

Figure 2. *Fitted Gamma and Normal distributions for reaction times across all participants. The Gamma distribution models a continuous distribution with two parameters (shape and rate) which is often used to model wait times and other phenomena that are always positive and skewed. When the shape parameter is > 1, the distribution is positively skewed. The Normal distribution is symmetric and models a continuous distribution with two parameters (mean and standard deviation).*

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 = -10149.80; Gamma = -10013.54. 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 experience type were not independent when excluding trials where participants responded 'I don't know' to the task-relevance question (χ^2 = 76.72, df = 4, p < .001).

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 %

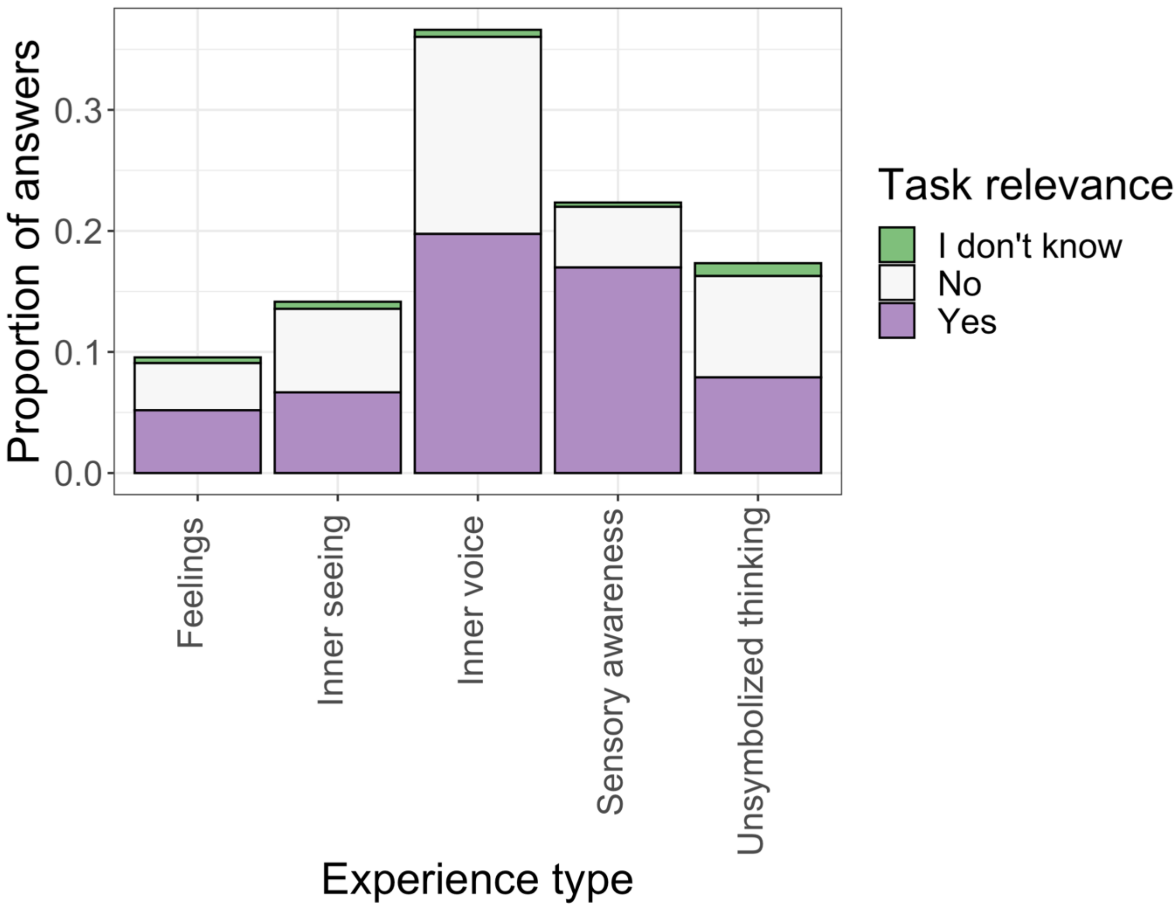

Figure 3. *Visualization of reported types of inner experience and whether they were relevant to the task or not across all prompts (eight per participant). The difference in proportions between task-relevant/task-irrelevant thoughts was significant. The main difference seems to be sensory awareness, which is reported more frequently for task-relevant thought.*

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.

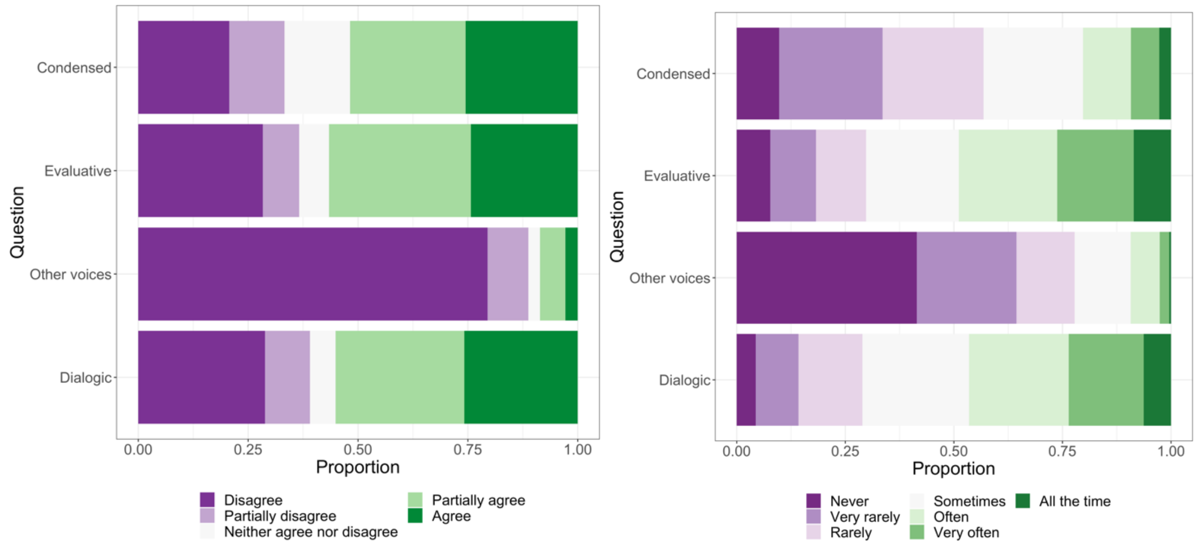

Figure 4. *On the left are the answers to the inner speech questions in the present experiment. On the right are the answers extracted from VISQ-R (Alderson-Day et al., 2018). Note that we had a five-point Likert scale while Alderson-Day et al. had a seven-point Likert scale.*

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 (β = -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 ($\beta = 0.01$, $SE = 0.005$, $p = 0.039$) 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 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 which simultaneously modelled 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.

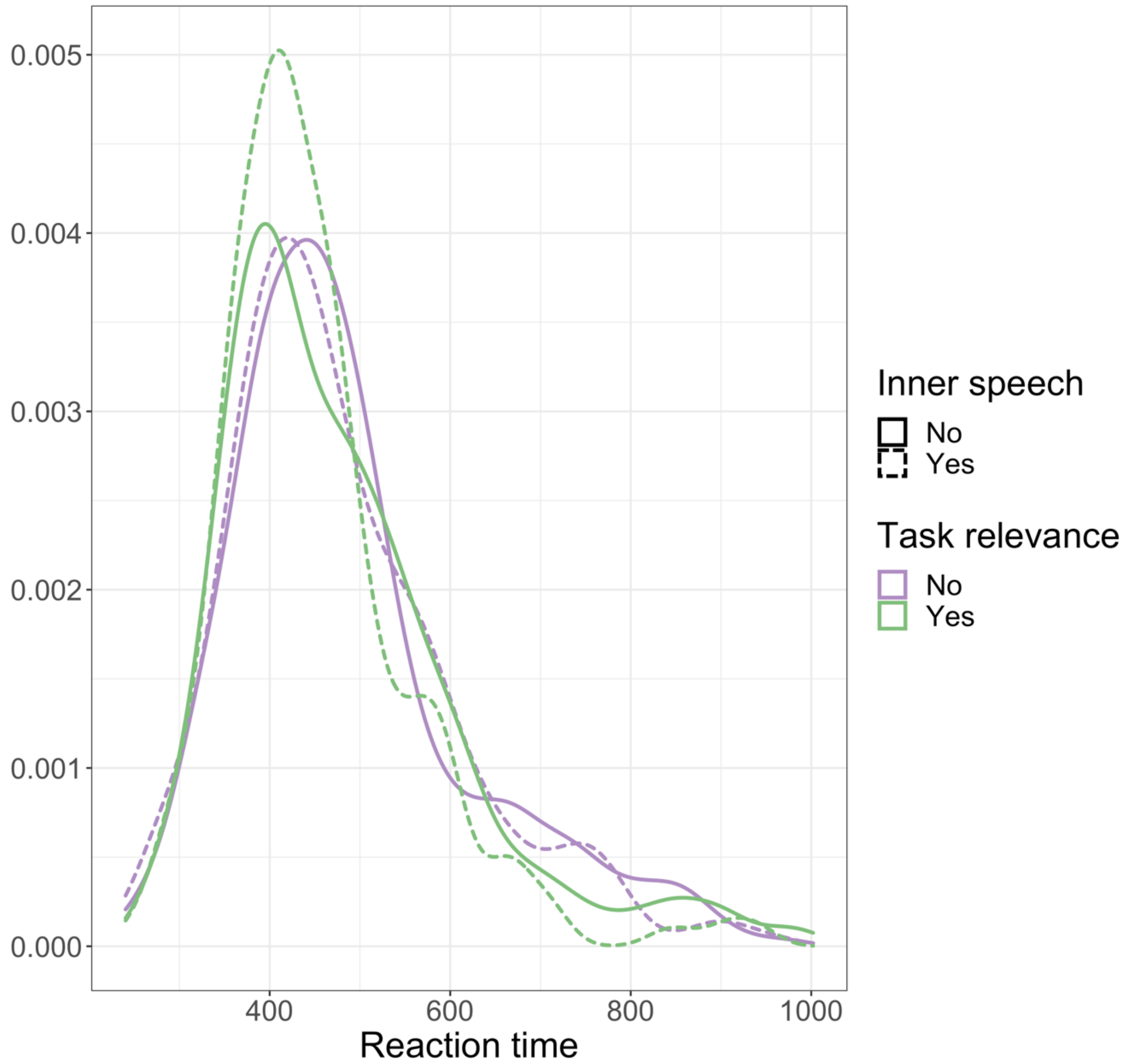

Figure 5. *Visualization of the densities of reaction times in the four combinations of task relevance and inner speech. Note that the mode of task relevant trials is lower than those of task-irrelevant trials. Also note that the*

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 modelling to investigate this.

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 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.

Priors. The prior for the difference in mode was modelled 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 ($\frac{1}{\sqrt{s^2}}$) was modelled 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 10000 iterations (first 5000 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 14000. The overall difference in log precision was 0.12 (95 % CI: -0.03 to 0.27).

The Rhat was 1.001, and the effective sample size was 15000. See Figure 6 below 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.

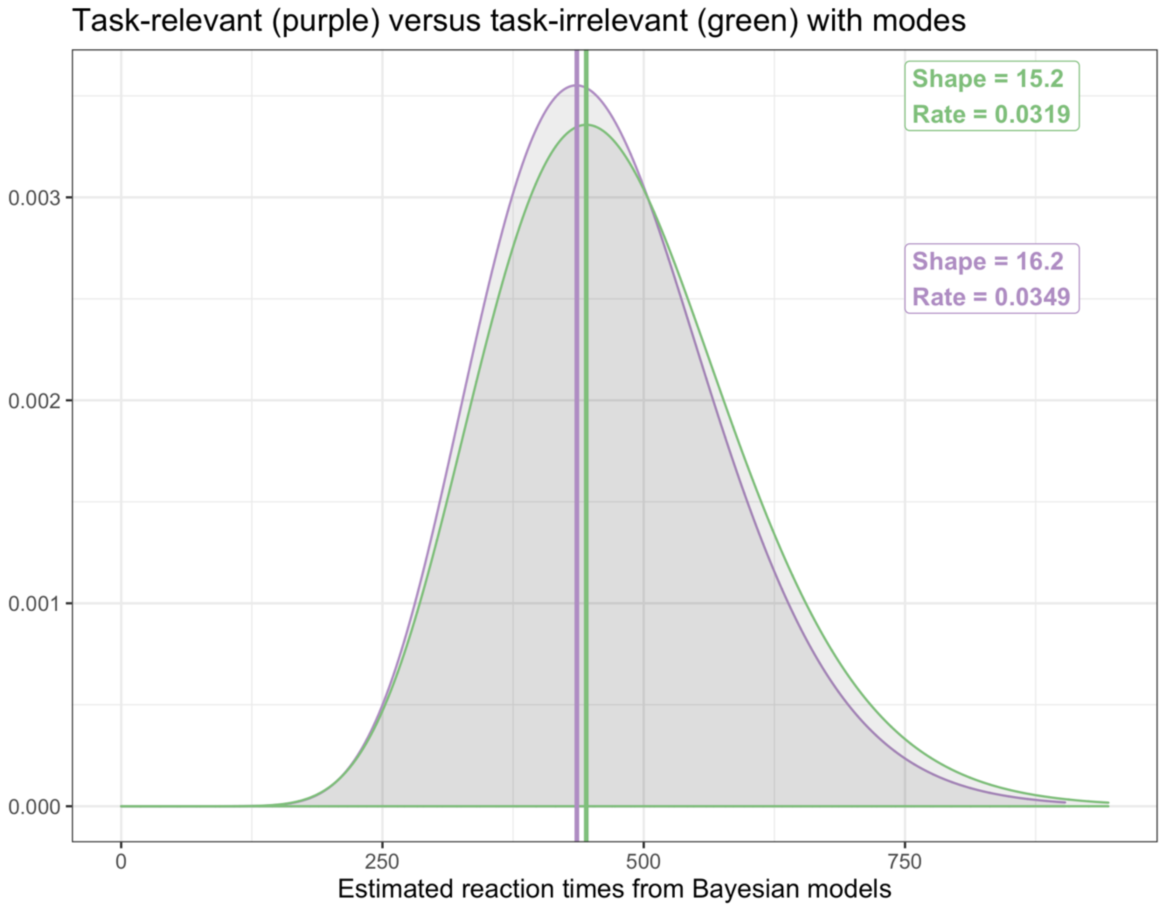

Figure 6. *Posterior estimates of the reaction time Gamma distributions on task-relevant (purple) and task-irrelevant (green). Vertical lines indicate modes.*

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 10000 iterations (first 5000 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 to 35.81). The Rhat was 1.001, and the effective sample size was 15000. The overall difference in log precision was -0.42 (95 % CI: -0.59

to -0.24). The Rhat was 1.001, and the effective sample size was 15000. 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 and task-irrelevant trials. See Figure 7 below for estimated Gamma distributions on trials preceded by task-relevant inner speech and all other trials.

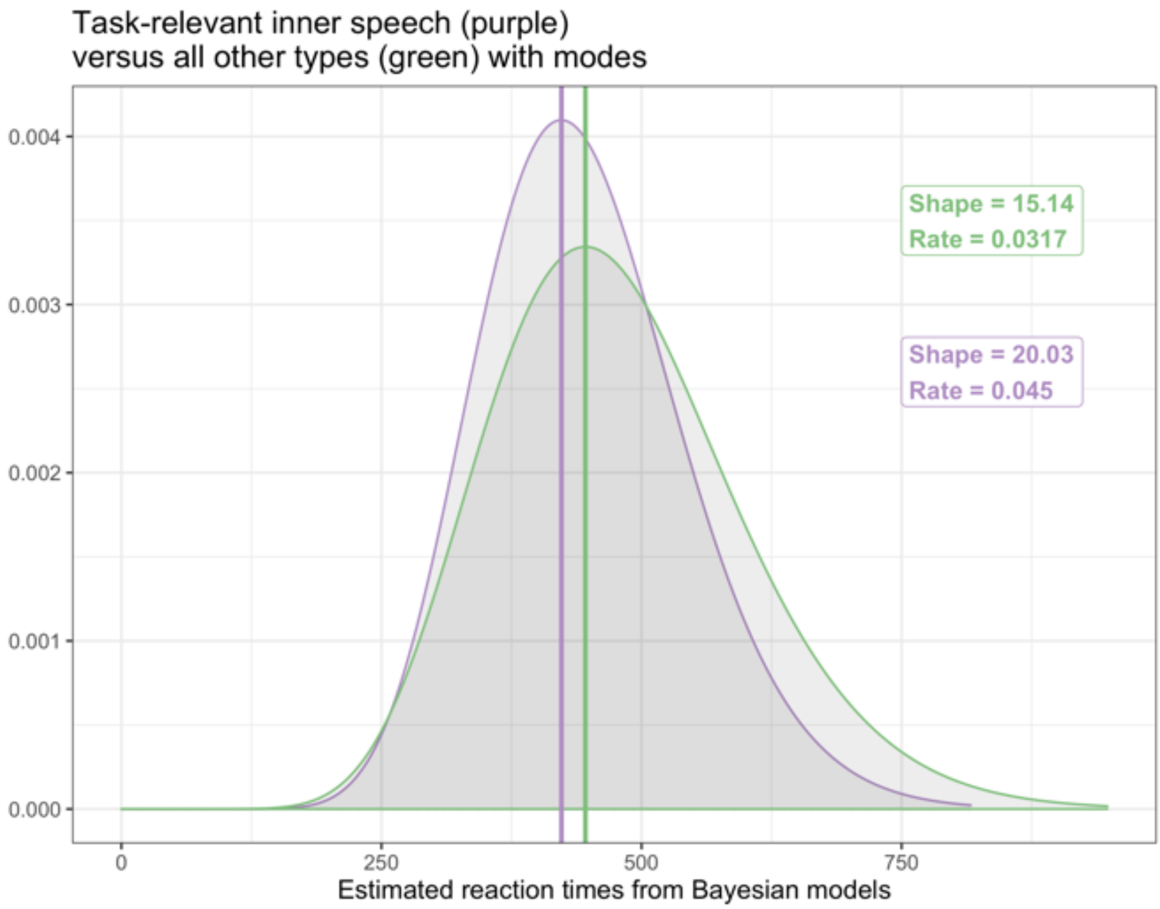

Figure 7. *Posterior estimates of the reaction time Gamma distributions on task-relevant inner speech trials (purple) and all other types of trials (green). Vertical lines indicate modes.*

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 10000 iterations (first 5000 discarded). The comparison was not preregistered, and the model specification can be accessed through the supplementary 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 to 30.61). The Rhat was 1.001, and the effective sample size was 15000. The overall difference in log precision was -0.39 (95 % CI: -0.58 to -0.19). The Rhat was 1.001, and the effective sample size was 15000. 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 below.

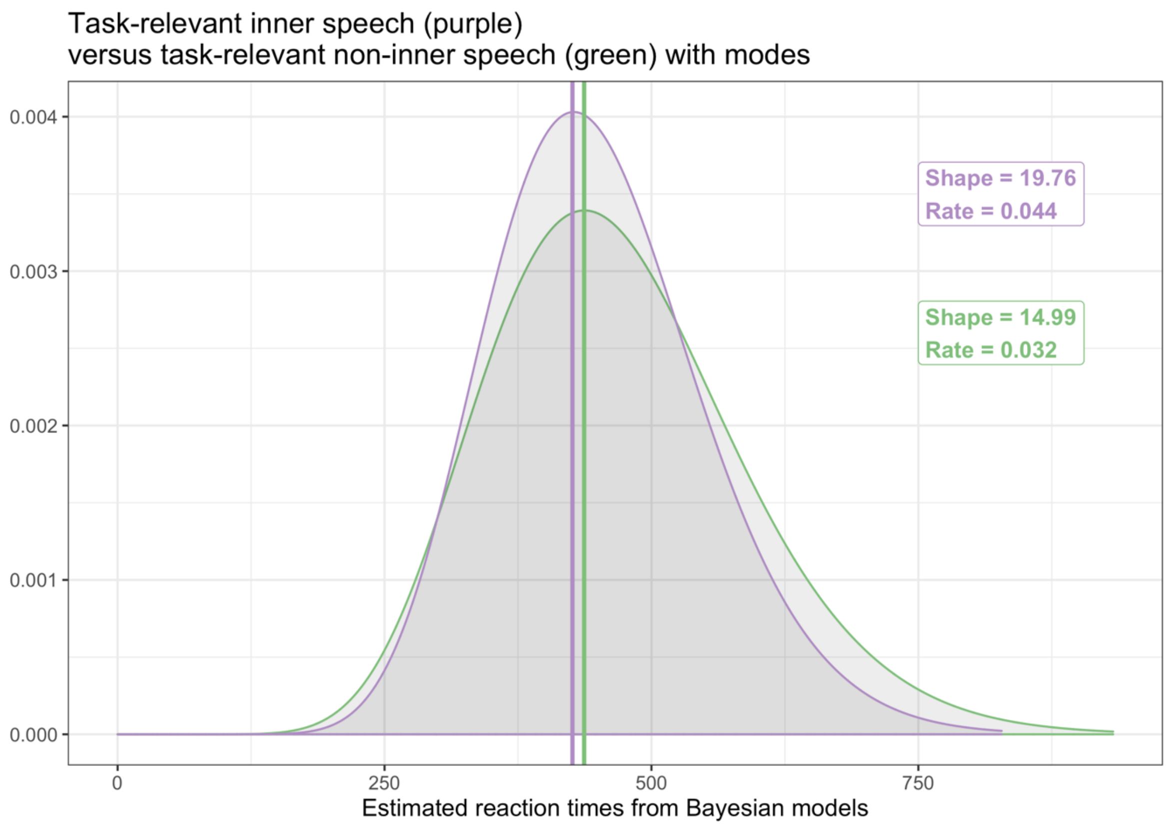

Figure 8. *Posterior estimates of the reaction time Gamma distributions on task-relevant inner speech trials (purple) and task-relevant non-inner speech trials (green). Vertical lines indicate modes.*

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 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).

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 inter-stimulus intervals was again randomized for each participant (30 seconds twice, 40 seconds, 50 seconds twice, 60 seconds twice, 70 seconds, 120 seconds twice, 150 seconds, 180 seconds). Because this increased the duration of the experiment from approximately 20 minutes to approximate 26 minutes, we also increased the compensation from £3 to £4. The data were collected in January 2022.

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 to 76)
Median time spent (range)	23 min 24 secs (18 min 30 secs to 69 min 59 secs)

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 > 1000 ms, 2.9 % of all trials). As predicted, the reaction time data were gamma rather than normally distributed (see Figure 9).

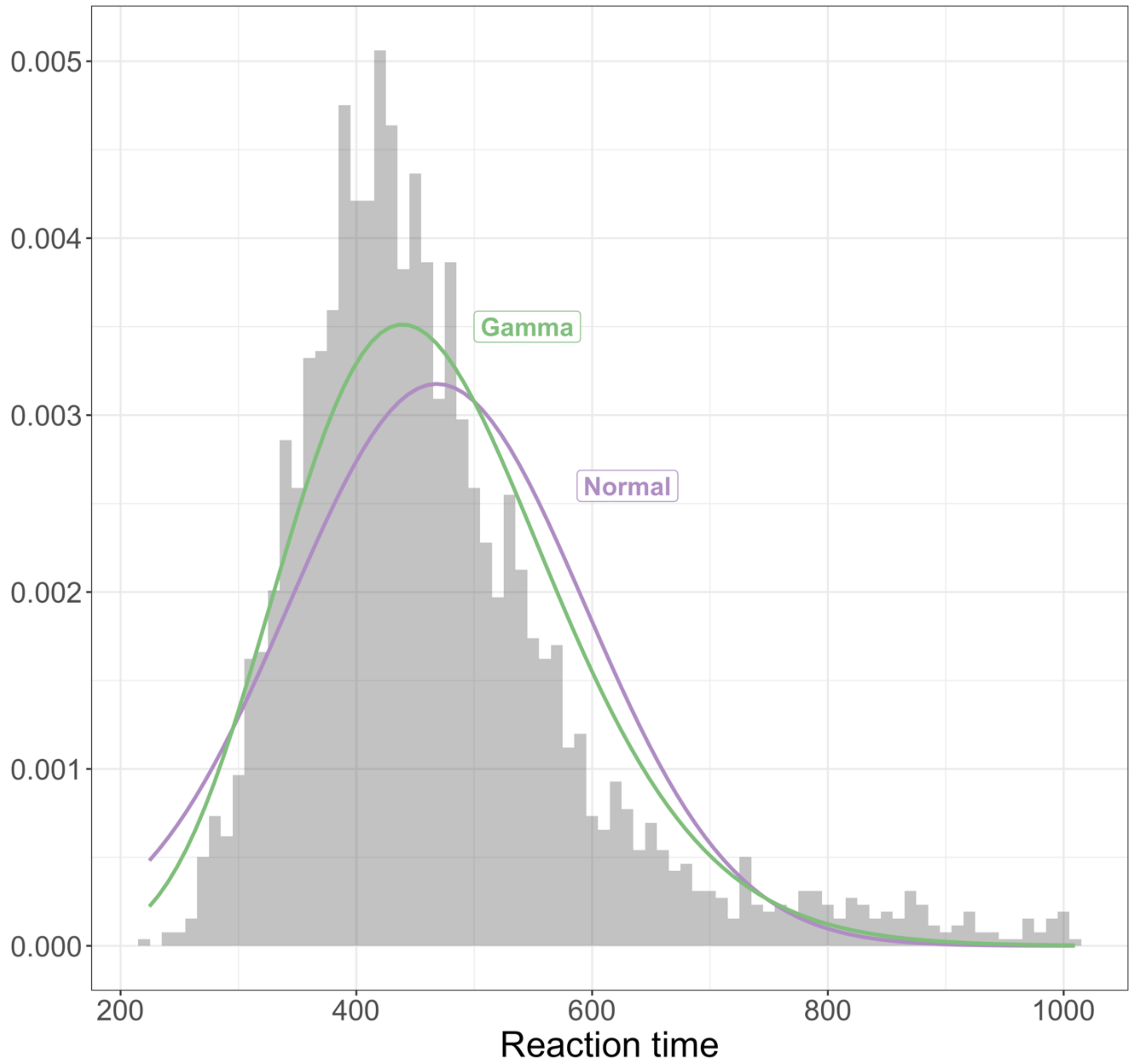

Figure 9. *Fitted Gamma and Normal distributions to all trials across participants. The Gamma distribution models a continuous distribution with two parameters (shape and rate) which is often used to model wait times and other phenomena that are always positive and skewed. When the shape parameter is > 1, the distribution is positively skewed. The Normal distribution is symmetric and models a continuous distribution with two parameters (mean and standard deviation).*

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 = -16180.03; Gamma = -15934.14. 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 remaining analyses.

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 (χ^2 = 112.45, df = 4, p < .001).

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 %

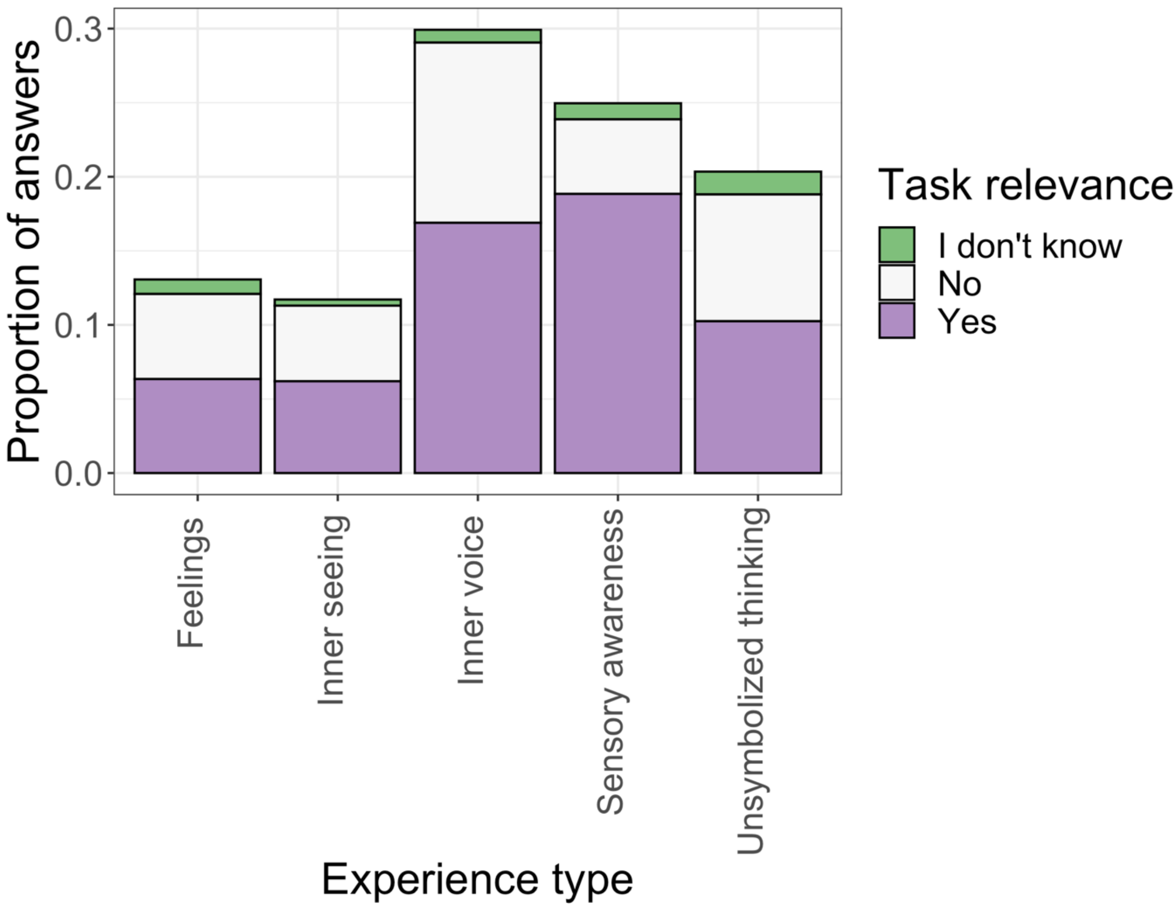

Figure 10. *Visualization of reported types of inner experience and whether they were relevant to the task or not across all prompts (twelve per participant).*

Inner speech questions

See Figure 11 below 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 the 24.1 % in the Experiment 1).

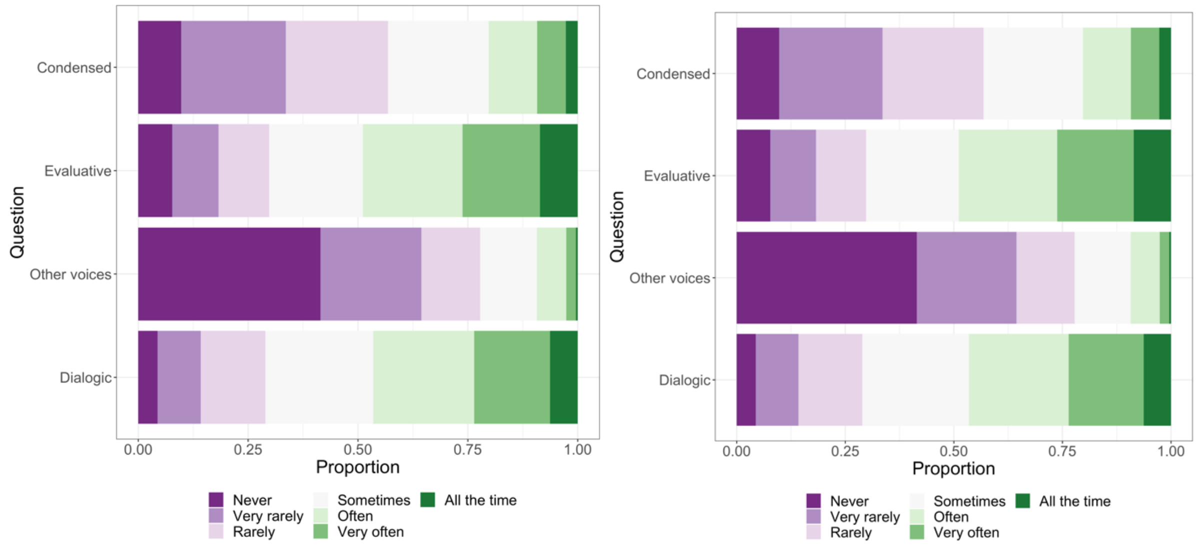

Figure 11. *On the left, we see the answers to the inner speech questions in the present experiment. On the right, we see the answers from VISQ-R (Alderson-Day et al., 2018). Note that we had a five-point Likert scale while Alderson-Day et al. had a seven-point Likert scale.*

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 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.

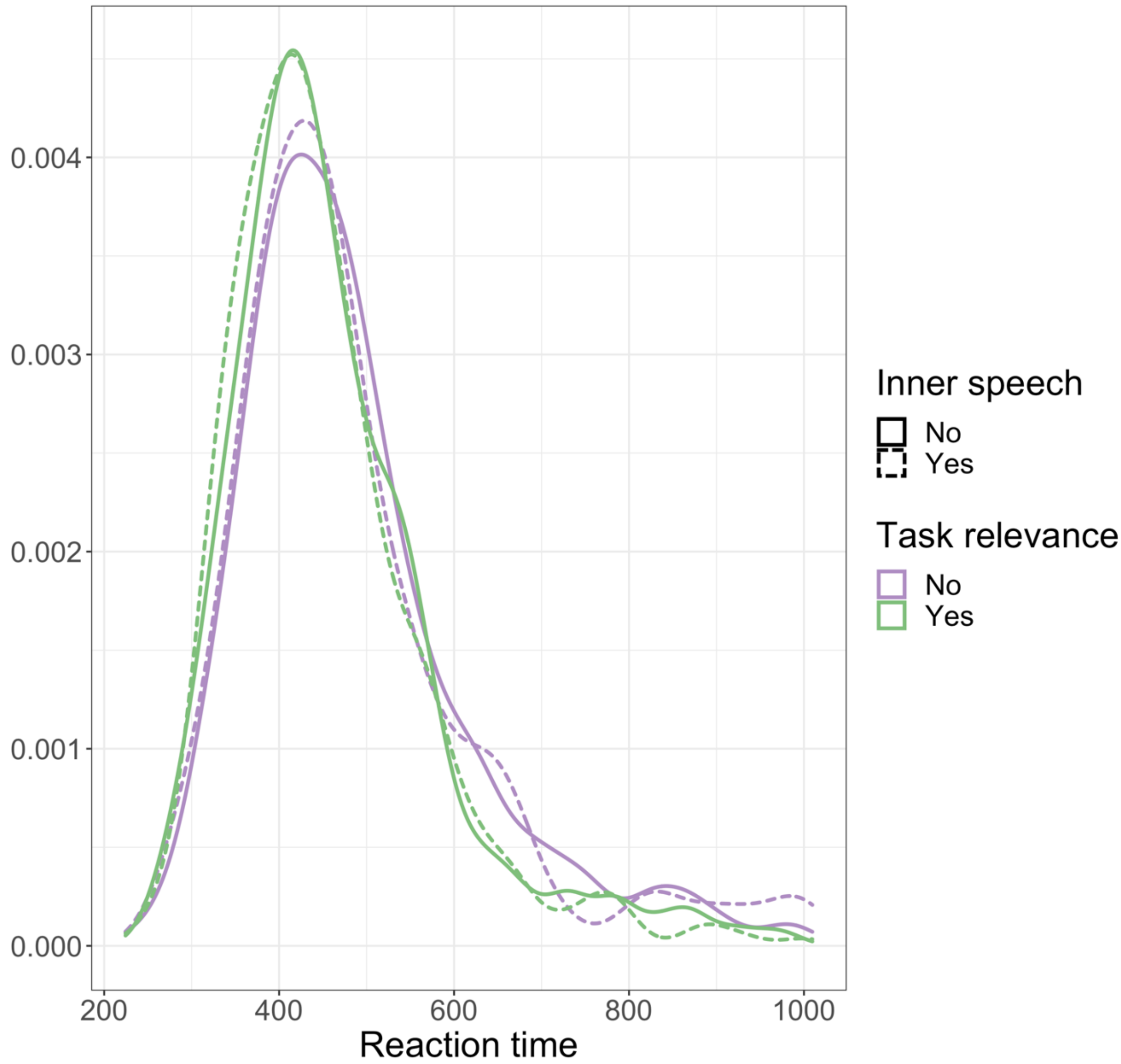

Figure 12. *Visualization of the densities of reaction times in the four combinations of task-relevance and inner speech. As in the original experiment, we used hierarchical Bayesian modelling to investigate the distributions.*

Priors

We used uninformative priors for both the difference in mode and standard deviation. The prior for the difference in mode was modelled as a Normal distribution with a mean of 0 and a standard deviation of 32 while the prior for the difference in log precision ($\frac{1}{\sqrt{s^2}}$) was modelled 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 10000 iterations (first 5000 discarded). The overall difference in mode between task-relevant and task-irrelevant trials was 18.3 ms (95 % CI: 8.29 to 27.84). The Rhat was 1.001, and the effective sample size was 14000. The overall difference in log precision was -0.28 (95 % CI: -0.4 to -0.16). The Rhat was 1.001, and the effective sample size was 15000. See Figure 13 below 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 (95 % CI: 0.77 to 23.98), and the overall difference in log precision was -0.25 (95 % CI: -0.4 to -0.10).

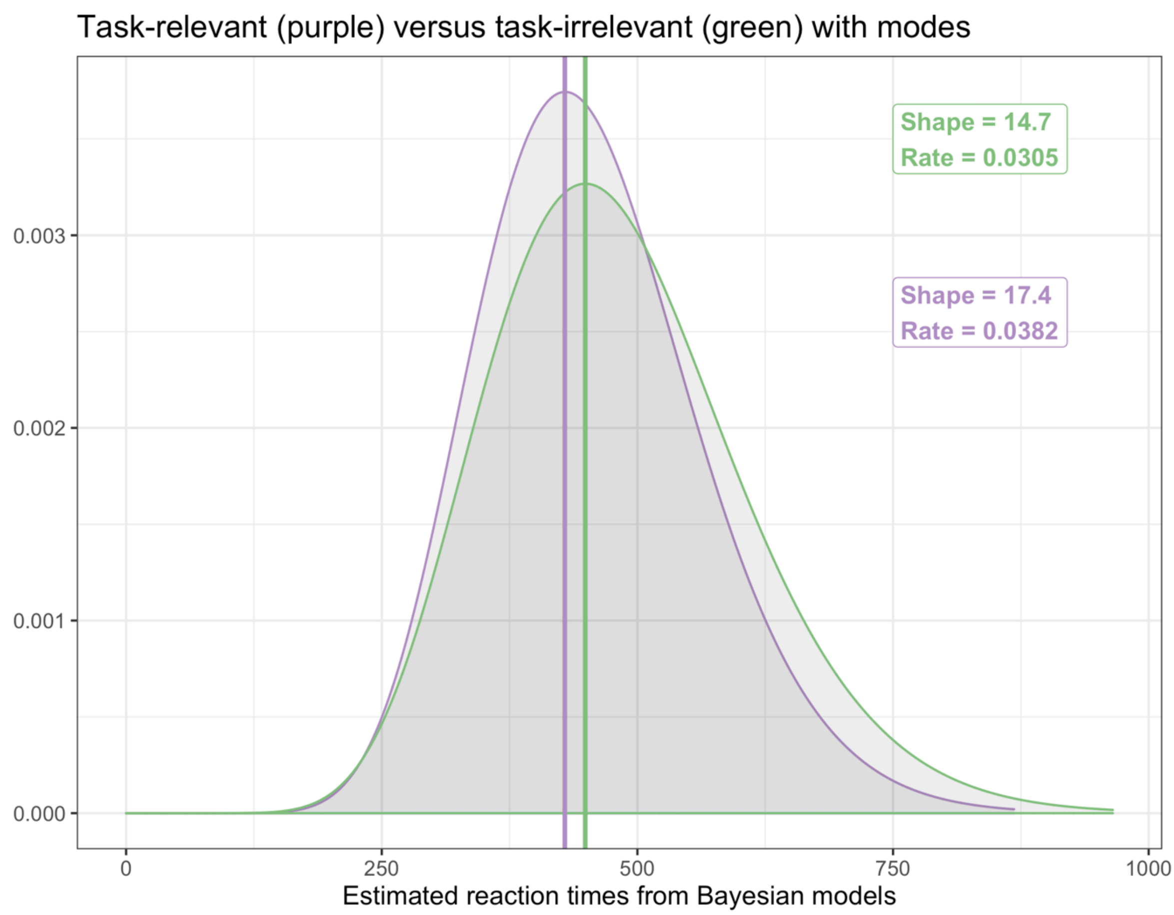

Figure 13. *Posterior estimates of the reaction time Gamma distributions on task-relevant trials (purple) and all task-irrelevant trials (green). Vertical lines indicate modes.*

Task-relevant inner speech against all other trials

For this model, we had three chains and 10000 iterations (first 5000 discarded). The overall difference in mode between task-relevant inner speech trials and all other trials was 18.08 ms (95 % CI: 7.51 to 28.57). The Rhat was 1.001, and the effective sample size was 15000. The overall difference in log precision was -0.3 (95 % CI: -0.45 to -0.15). The Rhat was 1.001, and the effective sample size was 3300. 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 below for estimated Gamma distributions on trials preceded by task-relevant inner speech and all other trials.

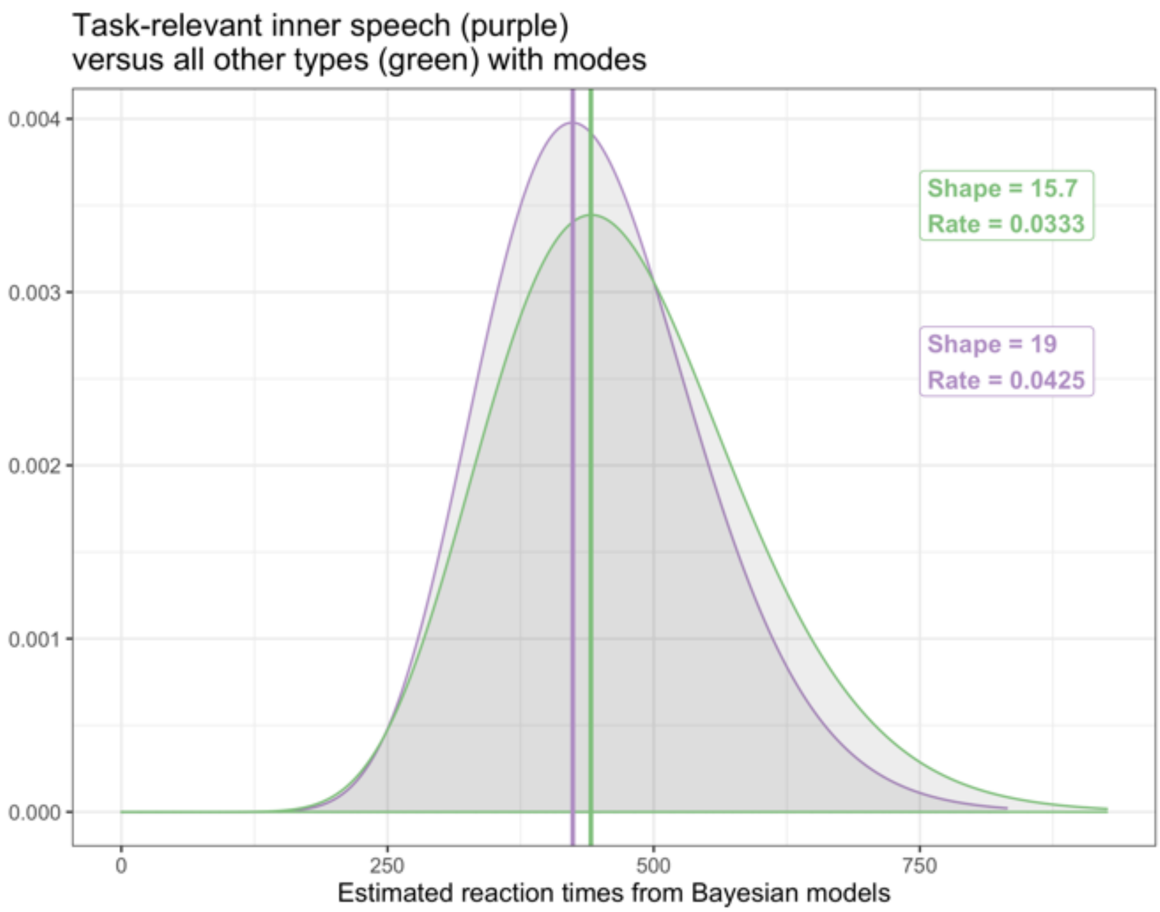

Figure 14. *Posterior estimates of the reaction time Gamma distributions on task-relevant inner speech trials (purple) and all other types of trials (green). Vertical lines indicate modes.*

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 10000 iterations (first 5000 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 3100. The overall difference in log precision was -0.21 (95 % CI: -0.37 to -0.05). The Rhat was 1.001, and the effective sample size was 4300. 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 below for estimated Gamma distributions on trials preceded by task-relevant inner speech and task-relevant non-inner speech trials.

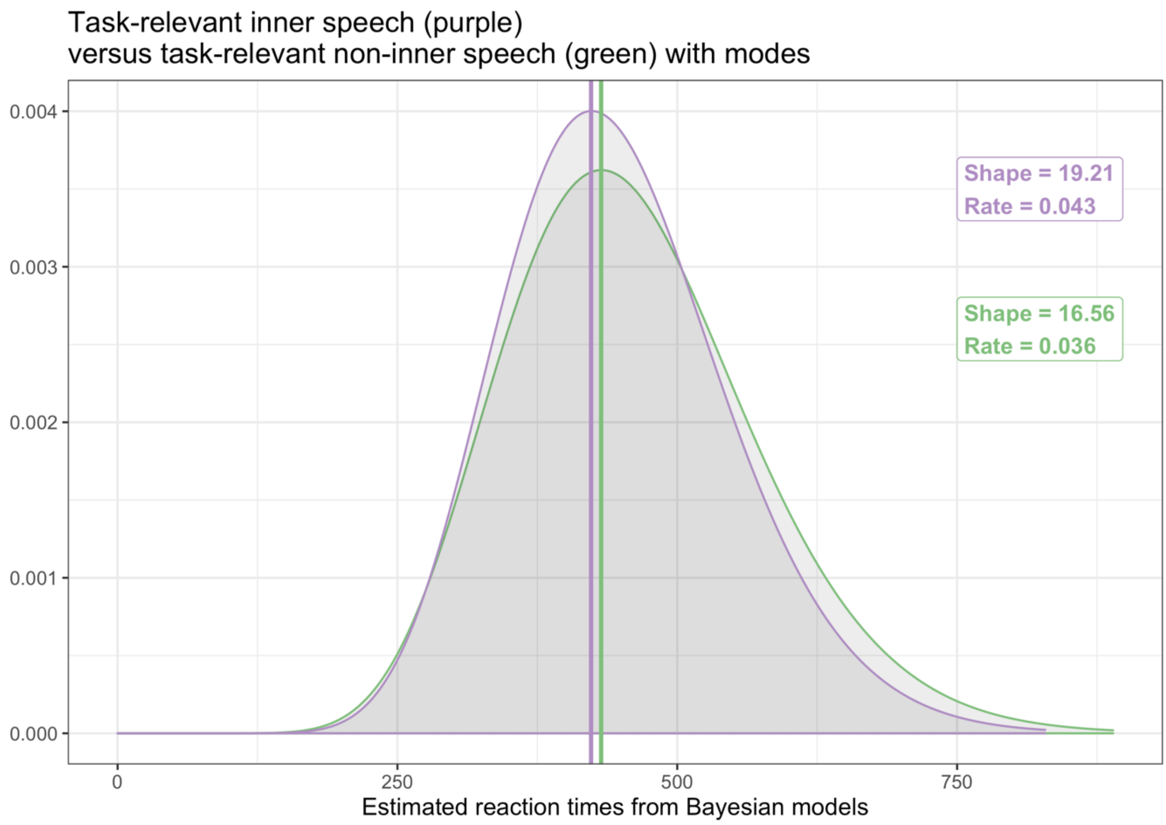

Figure 15. *Posterior estimates of the reaction time Gamma distributions on task-relevant inner speech trials (purple) and task-relevant non-inner speech trials (green). Vertical lines indicate modes.*

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: 1) Humans can use their inner voice for focused control of attention in tedious situations, 2) This use of inner voice can decrease reaction times and reduce variability in responses.

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*.

Modelling 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 command.

Task relevance presumably reduce 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 conceivably 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., 2021). 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: 1) Inner speech is a tool that can be used more or less productively, 2) 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.

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

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 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.

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

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.

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*, 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.

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. (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. (2021). *Mind over Body: Investigating Cognitive Control of Cycling Performance with Dual-Task Interference* [Conference poster]. 43rd Annual Meeting of the Cognitive Science Society. https://nedergaard.cogsci.au.dk/Nedergaard_Wallentin_210729.pdf

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.*

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*. 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.

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*, 287–290.

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-talk in competitive tennis. *Journal of Sport and Exercise Psychology*, *22*(4), 345–356.

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), 25.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 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