AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001316

Listening Speaks to Our Intuition While Reading Promotes Analytic Thought

Janet Geipel and Boaz Keysar
Department of Psychology, University of Chicago

It is widely assumed that thinking is independent of language modality because an argument is either logically valid or invalid regardless of whether we read or hear it. This is taken for granted in areas such as psychology, medicine, and the law. Contrary to this assumption, we demonstrate that thinking from spoken information leads to more intuitive performance compared with thinking from written information. Consequently, we propose that people think more intuitively in the spoken modality and more analytically in the written modality. This effect was robust in five experiments (N = 1,243), across a wide range of thinking tasks, from simple trivia questions to complex syllogisms, and it generalized across two different languages, English and Chinese. We show that this effect is consistent with neuroscientific findings and propose that modality dependence could result from how language modalities emerge in development and are used over time. This finding sheds new light on the way language influences thought and has important implications for research that relies on linguistic materials and for domains where thinking and reasoning are central such as law, medicine, and business.

Keywords: thinking, language, modality, intuition, analysis

Supplemental materials: https://doi.org/10.1037/xge0001316.supp

The ability to communicate and the ability to think are fundamental human skills. We think based on information that is communicated to us, and we communicate our conclusions using language. Theories of thinking are typically concerned with knowledge representation (Cheng & Holyoak, 1985) and with the rules and procedures that are performed on these representations (e.g., Braine, 1998; Rips, 1994). This is true regardless of whether a theory assumes a propositional representation (e.g., Rips, 1983) or an analogical one (Ford & Johnson-Laird, 1985; Johnson-Laird, 2010). In general, research has implicitly assumed that thought is influenced by the informational content, and not by the language modality through which the content is communicated.

Indeed, thinking *should* be modality independent because it is about the content of information. Consider the argument: "John is taller than Mark, and Mark is taller than Dave. Therefore, John is taller than Dave." The conclusion logically follows from the premises and the rule of transitivity. Because transitivity is conveyed via meaning, using transitivity should not depend on whether we read this argument or hear it. This is so self-evident that it is rarely stated and has been tacitly assumed in philosophy, logic, law, and psychology. Here we investigate the psychological validity of this assumption.

Thinking

Dual process models of thinking assume that two qualitatively different types of mental processes determine the way we think (Epstein, 1994; Kahneman & Frederick, 2002; Sloman, 1996, 2002). One is often characterized as more automatic and intuitive, and the other as more controlled and analytic (Hammond, 1996; Schneider & Shiffrin, 1977). For example, the default-interventionist account posits that intuitive processing delivers an initial response, which may or may not be monitored and corrected by subsequent analytic information processing (Evans, 2006; Kahneman, 2003). The parallel-competitive account instead holds that the intuitive and analytic processing routes operate simultaneously, but intuitive processing frequently forms the final response (Sloman, 1996, 2002).

Although dual process models have been highly influential, they are not universally accepted. A hybrid model suggests that logical intuitive processes could be responsible for certain cases that default-interventionist models have associated with analytic processes (De Neys, 2006). According to the hybrid model, there are two types of intuition: "heuristic" intuition

Janet Geipel https://orcid.org/0000-0003-1957-6213

This research was supported in part by the National Science Foundation (1520074) and the University of Chicago Center for International Social Science Research (CISSR). The authors thank Constantinos Hadjichristidis, Leigh H. Grant, Veronica Vazquez-Olivier, Zeynep Aslan, and Luca Surian for valuable comments on an earlier version of this article, as well as Yarra Elmasry and Lee Dong for editing. Both authors contributed equally. All data are available on the Open Science Framework, see https://osf.io/wyqh6/?view_only=63c7efa4e0e840e59b3ff9f5f67569dc (Geipel & Keysar, 2021).

Correspondence concerning this article should be addressed to Janet Geipel, who is now at Department of Management, University of Exeter Business School, Rennes Drive, Exeter EX4 4PU, United Kingdom. Email: j.geipel@exeter.ac.uk

that hampers performance on tricky logical problems and "logical" intuition that facilitates it. Furthermore, single process models argue against a qualitative difference between intuitive and analytic thinking processes, and instead view them as opposite ends of a continuum (e.g., Keren & Schul, 2009; Kruglanski, 2013; Kruglanski & Gigerenzer, 2011; Osman, 2004). Some suggest that the continuum reflects how accessible a thought is, and that the response to a problem depends on resources and motivation, which may influence such accessibility (see Kruglanski & Gigerenzer, 2011).

In what follows we assume that thinking can be relatively more intuitive or relatively more analytic, without making additional assumptions about the nature and interaction between these modes of thinking. That is, we do not associate our proposal to a particular dual process model (default-interventionist, parallel, or hybrid) nor does our account depend on whether thinking involves categorically distinct types of processes or a continuum. We propose that the language modality in which such problems are transmitted systematically influences thinking performance by differentially cuing heuristic intuition. Specifically, we hypothesize that spoken problems favor heuristic intuition and that written problems rely more on analytic processes.

Modality and Thinking

We reason that modality might influence thinking performance because reading and hearing language might differentially activate intuitive and analytic processing. This idea is motivated by neuroscientific research demonstrating differential neural responses when people read narratives as opposed to when they hear them (Michael et al., 2001; Regev et al., 2013). These differential patterns of neural activation are not simply related to low-level processing of sensory information, but to comprehending meaningful linguistic information. Specifically, spoken narratives elicit reliable activation in the left anterior dorsolateral prefrontal cortex, whereas written narratives elicit reliable activation in the left posterior dorsolateral prefrontal cortex. It has been suggested that such double dissociation implies the existence of distinct control processes for the two modalities, although the nature of the differences has not been specified (Regev et al., 2013).

Furthermore, it has been proposed that cognitive processes engaged in reading are more complex than those involved in listening (Liberman, 1989; Margolin et al., 1982; Rayner et al., 2012), and that reading places relatively more demands on cognitive processes that lead people to control, regulate, and maintain resources (Daneman & Merikle, 1996; Pimperton & Nation, 2010). This is supported by research showing that reading stories compared with listening to them results in fewer task-unrelated thoughts because cognitive resources are more occupied and people exert more cognitive control when reading (Kopp & D'Mello, 2016; Varao Sousa et al., 2013). It is possible that reading requires people to exert more cognitive control than listening and that as a result the act of reading may prompt more analytic thinking.[1]

The development of language and its use over time provide further motivations for the idea that the two modalities might differentially influence thinking performance. The acquisition of spoken language is spontaneous and effortless. The great majority of humans acquire spoken language by mere exposure to it (Liberman & Whalen, 2000; Pinker, 1994; Rayner et al., 2001; Rayner & Clifton, 2009). In contrast, the ability to read emerges later in development and requires intense formal instruction and practice (Rayner et al., 2001). Without instruction, many people never become literate (Liberman, 1989; Liberman & Whalen, 2000; Rayner et al., 2001; Rayner & Clifton, 2009). Indeed, 14% of the world population is illiterate (Roser & Ortiz-Ospina, 2018). This difference between the acquisition of spoken and written language has led some scholars to consider spoken language a human "instinct," and learning to read an intellectual achievement (Liberman, 1992; Liberman & Whalen, 2000; Musso et al., 2003; Pinker, 1994; Tomasello & Vaish, 2013). The later and relatively effortful acquisition of reading skills suggests a more deliberate process, whereas the early and relatively spontaneous acquisition of spoken language suggests a more intuitive process.

Furthermore, once spoken and written language are mastered, they are routinely used in different contexts. Spoken language is often used in casual settings and informal exchanges. In contrast, written language is frequently used in formal contexts such as school or work settings and in official documents. Written language also involves more formal language (Cunningham & Stanovich, 1998; Hayes & Ahrens, 1988), even when the exchange itself is not necessarily formal. It has even been argued that spoken language is a more primary means of communication than written language (McGregor, 2009). Such differences in the developmental trajectory and use of spoken and written language could promote relatively more heuristic intuitive responding when receiving information in spoken form and relatively more analytic responding when receiving information in written form.

There is evidence from other domains that performance on thinking problems depends on task, situation, and mindset. For example, it is well documented that individualistic and collectivist cultural mindsets can influence perception (Nisbett et al., 2001; Varnum et al., 2010) and problem solving (Arieli & Sagiv, 2018). For instance, Arieli and Sagiv (2018) studied bilingual-bicultural individuals and found differences in problem-solving performance depending on which cultural mindset was primed through the language used. Other evidence suggests that context can prime the use of more intuitive or more analytic thinking. For example, instructing participants to draw a picture of their current emotional state compared with solving mathematical problems influenced the quality of their subsequent decisions (Usher et al., 2011). Similarly, answering questions that require calculations primed a more analytic mindset compared with answering questions about feelings (Hsee & Rottenstreich, 2004). Finally, inducing a more mindful mindset reduced susceptibility to cognitive biases by increasing analytic responding (Maymin & Langer, 2021).

We propose that language modality might systematically influence the way people think. Because spoken language is spontaneously acquired early in development, and because it is routinely used in more informal contexts, it might make heuristic intuition relatively more accessible. On the other hand, because reading is more effortful and learned later in development and because written language is used in more formal contexts, it might favor relatively more analytic thinking. Our theory predicts that thinking

[1] We thank an anonymous reviewer for suggesting this alternative account.

based on spoken problems would result in comparatively more heuristic, intuitive responses while thinking based on written problems would result in comparatively more analytic responses. The experiments we report provide evidence for this modality dependence theory.

Design and Logic of the Experiments

In five experiments, we examined whether listening prompts relatively more intuitive responding and reading prompts relatively more analytic responding. We tested problems that involve a conflict between a heuristic, intuitive solution that for most people is readily accessible and a solution that for most people is less accessible and requires more analytic thought. We predicted that such conflict problems should yield relatively more heuristic responses when heard and relatively more analytic responses when read.

Consider the following problem that involves such a conflict "If you are running a race and you pass the person in second place, what place are you in?" (Thomson & Oppenheimer, 2016). Heuristic intuition suggests that you are now in first place but the more analytic response is that you are in second place (Oldrati et al., 2016; Travers et al., 2016; but see Bago & De Neys, 2017). All our experiments test verbal problems that include such a conflict between the answer supported by more heuristic intuition and by more analytical thought. If our modality dependence theory is correct, then written problems would yield relatively more analytical responses while spoken problems would cue relatively more intuitive responses.

Study 1 investigates the modality dependence theory with a simple task that requires the identification of semantic anomalies. Studies 2, 3, and 4 investigate it with insight problems like the race problem described above. Finally, Study 5 examines the effect of modality on deductive reasoning. All studies tested native speakers of English and used English materials except for Study 3 that examined native speakers of Chinese and used materials in Chinese. Its purpose was to examine whether the theory extends to other languages and to a non-Western, educated, industrialized, rich, and democratic (WEIRD) population (Henrich et al., 2010).

Methodological Notes

Presentation Format

In all the studies, we aimed to test how language modality influences thinking performance in a relatively ecologically valid way while preserving experimental control. In the written condition participants read each problem in a self-paced manner as this reflects a natural way of reading. In the spoken condition participants heard the problem in a natural pace. Following each problem, participants had to click on a button labeled "next" that directed them to a new page where they were asked to provide their response. Once participants clicked the button, the problem was no longer accessible to them. This ensured that in both modalities they could not go back and review the problem. We note variations in the presentation format in the Method of each experiment.

Response Format

Because our theory does not concern the response format, we elected to keep the response format the same across modalities and asked participants to type their response (or click in some cases) in both modality conditions. This choice also had the advantage of presenting a "switch" in both modality conditions. They switched from reading to typing, or from listening to typing.

Study 1

Study 1 investigates thinking performance that requires the identification of semantic anomalies by using what is known as the "Moses Illusion" (Erickson & Mattson, 1981). When asked "How many animals of each kind did Moses take on the Ark?" people typically rely on their intuition and respond "two." But if people instead reflect further they will realize that the very question is wrong (Fazio et al., 2015). It was Noah, not Moses, who gathered animals on the ark. Therefore, the question itself involves a semantic anomaly. The reason people show this illusion is that Moses and Noah are two biblical characters who are highly associated semantically with each other (Sanford & Sturt, 2002). Heuristic intuitive thinking relies on such associations and provides an incorrect answer. To notice the anomaly, one must suppress this erroneous intuition and access stored knowledge. Therefore, the Moses Illusion is a useful tool to investigate our proposal. If listening prompts relatively more heuristic intuitive thinking, then it should prompt more intuitive responses, which in this case is "two of each animal." In contrast, when such questions are read, people would be more likely to detect the semantic anomalies.

Method

The data of all our studies are available on the Open Science Framework, see https://osf.io/wyqh6/?view_only=63c7efa4e0e840e59b3ff9f5f67569dc (Geipel & Keysar, 2021). All materials are presented in the Supplemental Materials Method. The study prediction, study design, sample size, and analyses were preregistered on AsPredicted.org, see https://aspredicted.org/87j8g.pdf. The University of Chicago Institutional Review Board (IRB) approved the research of all studies.

Power Analysis

We conducted an a priori power analysis for a dependent t test (two-tailed) with the following estimates: $\alpha = .05$, power = .95, $d_{\text{Cohen}} = .40$ (based on piloting). This analysis revealed that a minimum of 84 participants was required for a within-participants design. As a precaution of possible data loss, we preregistered and requested 100 participants through Prolific (www.prolific.co).

Participants

Participants were native English-speaking U.S. residents and recruited online through Prolific (prolific.co). We received data from 107 participants of whom we excluded 18 (16.8%): 6.5% because they failed an audio check at the beginning of the study and 10.3% because they failed at least two out of four catch questions distributed across the study. We analyzed the data of the remaining 89 participants (42 women, 46 men, 1 other, M_{age} = 33.1 years, age range = 18 to 60 years).

Material and Procedure

Participants received 30 trivia questions, 15 in the written and 15 in the spoken modality. Thus, modality was manipulated within participants. In each modality condition, participants received five questions that were wrong, or semantically "distorted," and 10 that were undistorted control questions (see Supplemental Materials Method Table S1 for the full set of questions). The presentation order of the spoken and written questions was blocked, and the presentation order of the modalities was counterbalanced across participants. We also counterbalanced across participants the list of items in each modality condition so that each question appeared in both modalities across subjects. Furthermore, within each modality block, the presentation order of the 15 trivia questions was randomized separately for each participant.

Each trivia question had three response options: A blank box to type a response, a "Do not know" button, and a "Wrong" button. We instructed participants that they may or may not encounter questions that have something wrong with them, and which do not have an answer if taken literally (see online supplemental materials for exact wording of the instructions). Then we provided participants with an example of such a distorted question and explained that the question itself is wrong, and that whenever they encounter such a question they should answer "Wrong". In the spoken modality, the trivia questions were read aloud in a neutral tone by two male native English speakers with a standard American English accent.

Following the trivia questions, participants were asked "Overall, how would you rate your performance on this task?" (1 = *worse performance* to 101 = *best performance*), and then completed a short version of a Need for Cognition and Faith in Intuition scales (Cacioppo et al., 1984). The Need for Cognition scale included statements such as: "I try to avoid situations that require thinking in depth about something.,", "I prefer complex to simple problems." The Faith in Intuition scale included statements such as: "I can usually feel when a person is right or wrong even if I cannot explain how I know it.," and "When it comes to trusting people, I can usually rely on my 'gut feelings.'" (1 = *completely false* to 5 = *completely true*).

Subsequently, participants answered questions about the task such as "Overall, how involved did you feel in this task?," "Overall, how interesting was this task for you?," "Overall, how bored were you?," and "Overall, how much did you like this task?" (all on a slider scale, 1 = *not at all*, 101 = *very much*). Participants in the spoken condition were then asked "Overall, how well did you understand the speaker?" (1 = *not at all well*, 7 = *perfectly well*) and "Overall, how much did you like the voice of the speaker?" (1 = *not at all*, 7 = *very*; see Supplemental Materials Results Tables S1 to S3 for descriptive statistics). Finally, participants were asked to indicate their age, gender, and educational level (see Supplemental Materials Results Table S6 for descriptive statistics).

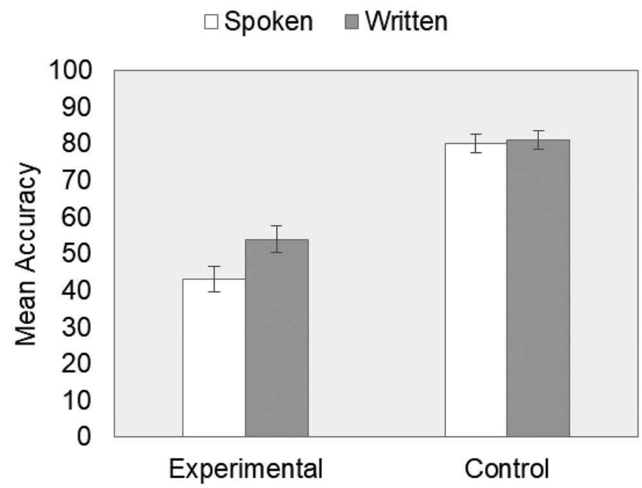
Results and Discussion

Distorted Questions

We calculated the mean rate of correct identification of the distorted questions for each participant (see Figure 1). Participants detected distortions significantly more often in the written than in the spoken condition (M_{Written} = .53, CI [.46, .61], *SD* = .34, M_{Spoken} = .43, 95% CI [.36, .49], *SD* = .32), $t(87) = 2.98$, $p = .004$, $d_{\text{Cohen}} = .32$ (Wilcoxon signed-ranks test: T = 1446.50, $z = 2.72$, $p = .006$, $d_{\text{Cohen}} = .39$).

Figure 1

Mean Accuracy by Language Modality and Question Type (Study 1)



Note. Error bars illustrate standard errors of the mean.

We submitted the mean rate of correct responses to a 2 (Order: spoken first, written first) × 2 (Speaker: male 1, male 2) × 2 (Modality: spoken, written) mixed-factor analysis of variance (ANOVA) with Order and Speaker as between-subjects factors and Modality as the within-subject factor. This analysis revealed a significant main effect of modality, $F(1, 84) = 8.76$, $p = .004$, $\eta_p^2 = .09$, which was qualified by a significant Modality × Order interaction, $F(1, 84) = 14.97$, $p < .001$, $\eta_p^2 = .15$. Pairwise tests, adjusted for multiple comparisons, showed that the modality effect was significant for participants who received the questions first in the spoken modality ($p = .001$), but not for those who first received the questions in the written modality ($p = .507$). However, there was no main effect of order, $F(1, 84) = .64$, $p = .428$, $\eta_p^2 < .01$, nor a main effect of speaker, $F(1, 84) = 1.79$, $p = .184$, $\eta_p^2 = .02$. There was also no significant Modality × Speaker interaction, $F(1, 84) = .99$, $p = .322$, $\eta_p^2 < .01$, nor was there a three-way Modality × Order × Speaker interaction, $F(1, 84) < .01$, $p = .927$, $\eta_p^2 < .01$. In general, participants were less able to detect the anomaly when hearing than when reading the problem, which suggests that they engaged heuristic intuitive processes more when listening.

We then analyzed the rate of intuitive answers. In the Moses question, this amounts to typing "two animals." Participants were significantly less likely to provide such intuitive yet wrong answers in the written than in the spoken condition (M_{Written} = .37, 95% CI [.31, .43], *SD* = .29, M_{Spoken} = .49, 95% CI [.43, .56], *SD* = .31), $t(87) = 3.46$, $p = .001$, $d_{\text{Cohen}} = .41$ (Wilcoxon signed-ranks test: $T = 573.00$, $z = 3.32$, $p = .001$).

Finally, we averaged the rate of the "Do not know" responses for each participant and found no difference between the modalities (M_{Written} = .09, 95% CI [.05, .13], *SD* = .19, M_{Spoken} = .08, 95% CI [.04, .12], *SD* = .18), $t(87) = -.69$, $p = .489$, $d_{\text{Cohen}} = .07$ (Wilcoxon signed-ranks test: $T = 396.50$, $z = 1.05$, $p = .295$).

Control Questions

The average rate of correct responses was comparable in the two modalities (M_{Written} = .81, 95% CI [.76, .86], SD = .23, M_{Spoken} = .80, 95% CI [.75, .85], SD = .24), $t(88)$ = .54, p = .589, d_{Cohen} = .05 (Wilcoxon signed-ranks test: T = 709.00, z = .06, p = .954). Similarly, the rate of the "Do not know" responses was comparable across modalities (M_{Written} = .16, 95% CI [.12, .21], SD = .23, M_{Spoken} = .17, 95% CI [.12, .21]), SD = .22, $t(88)$ = .12, p = .907, $d_{\text{Cohen}} < .01$ (Wilcoxon signed-ranks test: T = 683.00, z = .06, p = .956).

Next, we averaged the rate of "Wrong" responses for the control questions. Such a response would indicate that participants identified an anomaly when no anomaly was present. Such responses were equally rare in both modalities (M_{Written} = .03, 95% CI [.01, .04], SD = .06, M_{Spoken} = .04, 95% CI [.01, .06], SD = .12), $t(88)$ = .68, p = .496, d_{Cohen} = .10 (Wilcoxon signed-ranks test: T = 163.00, z = .01, p = .989).

Secondary Exploratory Results

We analyzed the results for the Need for Cognition, Faith in Intuition, self-rated performance in the reasoning task, and perceived task involvement. We found no support for a mediation or a moderation with any of these measures. We present a full report in the Supplemental Materials Results.

In summary, the results suggest that language modality systematically influences thinking performance in this task. When people respond to simple trivia questions, they are less likely to detect semantic anomalies when they hear them than when they read them. This performance difference suggests that the spoken modality makes heuristic intuitive thought more accessible than the written modality.

We found an order effect such that the performance advantage for the written modality was found when participants completed the spoken version first. All subsequent studies varied modality between-participants, which avoids order effects. One might suggest that the modality effect is due to a greater difficulty in comprehending spoken language because it is transient. Perhaps because listeners cannot go back and replay the spoken information, as opposed to being able to review the written information, they might experience more difficulty holding spoken information in memory. If this were the case, the performance deterioration in the spoken condition should also be observed for the undistorted control questions that were closely matched to the distorted experimental questions. Given that there was no modality difference in performance in the undistorted control questions, this finding argues against such an account.

Study 2

Study 2 extended the investigation to verbal riddles, such as the race problem presented in the introduction, known as the Cognitive Reflection test (CRT; Frederick, 2005; Thomson & Oppenheimer, 2016). These problems involve a conflict between a heuristic intuitive answer and a more analytic one. We purposefully chose verbal problems that do not require the application of complex mathematical and logic principles. We compared performance on such problems to performance on control problems where intuitive and analytic thinking do not provide conflicting answers ("What did Rudolph's nose do to help guide Santa's sleigh?"; Sirota et al., 2021). We expect that in the spoken modality people will rely more on heuristic intuitive thinking and would make more errors than in the written modality. However, it is possible that more errors in the spoken modality might result from factors other than an enhanced role for intuition. If this is the case, then performance should be worse in the spoken modality not just in the conflict problems but also in the non-conflict problems. Our theory predicts that thinking performance would be more intuitive in the spoken modality and less accurate than in the written modality, but only for the riddles that involve a conflict, not for the control ones.

Method

Power Analysis

We estimated the required number of participants by conducting an a priori power analysis for a one-way ANOVA using *G*Power* using the following estimates: α = .05, power = .95, f = .22 (based on piloting), number of groups = 2. This analysis revealed that a minimum of 272 participants was required. We recruited more participants to prevent a reduction in statistical power due to possible exclusions.

Participants

Participants were native English-speaking U.S. residents and were randomly assigned to the spoken or written condition. We recruited 353 participants (130 women, 216 men, 7 other, M_{age} = 38.6 years, age range = 18–88 years) from August 2018 to September 2018. Of these, 212 participants (40.8% women) were assigned to the spoken condition (female speaker n = 110, male speaker n = 102) and 141 participants (59.2% women) to the written condition. Given the overexposure of highly educated populations to reasoning problems such as the CRT (Haigh, 2016), we recruited participants from the downtown Center of the Decision Research of the University of Chicago, which attracts participants from a range of backgrounds.

Material and Procedure

Participants sat in a lab room in front of a computer screen with headphones and first rated their mood ("Overall, my mood is" on a scale ranging from −10 = *very unpleasant* to 10 = *very pleasant*). We measured participants' mood because studies have shown that mood can influence reasoning performance (e.g., Channon & Baker, 1994; Oaksford et al., 1996). Participants then received five verbal conflict problems and two control problems (see Supplemental Materials Method Table S2). As in Study 1, following each problem participants provided their answer on a separate page without the option to return to the actual problem. In the spoken modality, either a female or a male native English speaker with a standard American English accent read the problems aloud in a neutral tone. We randomized the problem order across participants. Participants provided their answer by typing it in a blank box on the computer. Once participants responded, they could continue to the next problem.

After answering the problems, participants received the following questions: "Overall, how involved did you feel in this task?" (slider scale, 1 = *not at all*, 101 = *very much*) and "Overall, how difficult was it for you to respond to the questions?" (1 = *not at all difficult*, 7 = *very difficult*; see Supplemental Materials Results Table S1 for descriptive statistics). Then participants answered a short form of the Need for Cognition and Faith in Intuition scales (Cacioppo et al., 1984), with each scale consisting of five statements (see Supplemental Materials Results Table S2 for descriptive statistics).

Participants in the spoken modality condition then rated the speaker ("Overall, how well did you understand the speaker?"; 1 = *not at all well*, 7 = *Perfectly well*, and "Overall, how much did you like the voice of the speaker?"; 1 = *not at all*, 7 = *very*; see Supplemental Materials Results Table S3 for descriptive statistics; see Supplemental Materials Results for exploratory analyses).

Finally, all participants reported their age, gender, political orientation, religious beliefs (1 = *not at all religious* to 7 = *very religious*), education level, and employment status (see Supplemental Materials Results Table S4 to S6 for descriptive statistics; see Supplemental Materials Results for exploratory analyses).

Results and Discussion

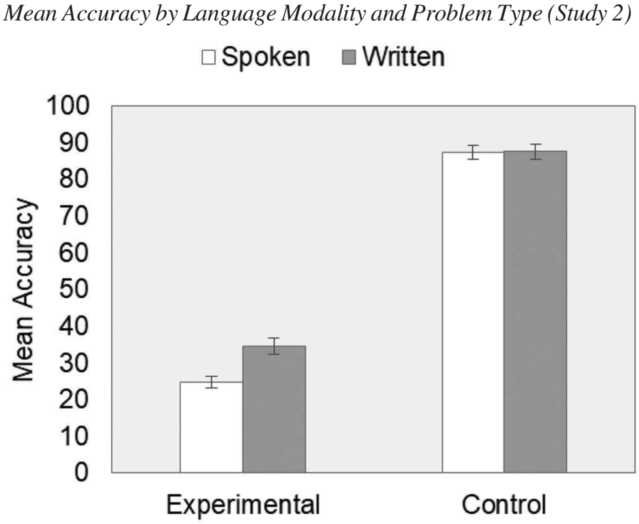
We first considered the speaker ratings and found that both speakers were understood equally well, Welch's $F(1, 202.89) = .30$, $p = .586$, $d_{\text{Cohen}} = .07$, and were equally liked, Welch's $F(1, 195.12) = 2.86$, $p = .092$, $d_{\text{Cohen}} = .24$ (see Supplemental Materials Results Table S3 for statistics). Because speaker identity did not influence reasoning accuracy, Welch's $F(1, 210) = .05$, $p = .827$, $d_{\text{Cohen}} = .03$, we combined results into a single spoken condition.

Conflict Problems

We computed the average accuracy rate over the five conflict problems for each participant (see Figure 2). We tested for differences between the conditions using the robust Welch's *F* test because the assumption of homogeneity of variance was violated. As predicted by our theory, mean accuracy was significantly higher in the written than in the spoken condition ($M_{\text{Written}} = .35$, 95% CI [.30, .39], $SD = .27$, $M_{\text{Spoken}} = .25$, 95% CI [.22, .28], $SD = .23$), Welch's $F(1, 263.05) = 12.45$, $p < .001$, $d_{\text{Cohen}} = .40$ (Mann–Whitney $U = 11{,}958.00$, $p = .001$, $d_{\text{Cohen}} = .34$).

Figure 2

Mean Accuracy by Language Modality and Problem Type (Study 2)



Note. Error bars illustrate standard errors of the mean.

Control Problems

We then averaged the number of correct responses across the control problems and found no difference in accuracy rate between the modality conditions ($M_{\text{Written}} = .88$, 95% CI [.83, .92], $SD = .25$, $M_{\text{Spoken}} = .87$, 95% CI [.84, .91], $SD = .26$), Welch's $F(1, 306.56) = .01$, $p = .941$, $d_{\text{Cohen}} = .04$ ($U = 14{,}828.00$, $p = .972$, $d_{\text{Cohen}} = .01$).

In summary, the results show that the spoken modality promotes more intuitive responding than the written modality. Given that we found no differences for control problems, these results support our modality dependence theory according to which the spoken modality cues comparatively more heuristic intuition than the written modality. This generalizes the language modality effect on thinking performance from the relatively simple verbal task of detecting semantic anomalies to the relatively more complex task of solving verbal insight problems.

Study 3

Writing systems vary between languages. We considered it essential to evaluate whether the language modality effect we discovered is specific to English or can be generalized to other languages. We chose Chinese for two reasons. While English is an Indo-European language, Chinese belongs to the Sino-Tibetan language group. The two languages use different writing systems with fundamentally different mapping between the sound of the word and the way it is written. English uses a segmental, alphabet system while Chinese uses a logographic system. If modality has the same effect in Chinese, it is reasonable to conclude that its impact generalizes to other languages beyond English. In addition to generalizing to a different language, the study also generalizes to a population that is not "WEIRD." Finally, different from Study 2, Study 3 used the same number of conflict and control problems so that we can compare them directly.

Method

Power Analysis

We estimated the number of participants by conducting an a priori power analysis for a mixed-factor ANOVA using an uncertainty bias correction (Anderson et al., 2017). The following estimates were used (based on Study 2): $\alpha = .05$, power = .80, $F = 17.19$, $N = 386$, number of between-subjects factor = 2, number of within-subject factor = 2. This analysis revealed that we needed a minimum of 296 participants. We recruited more participants to prevent reduction in statistical power due to possible exclusions.

Participants

Participants were native Mandarin Chinese speakers from Beijing and they were randomly assigned to the spoken or written condition. We recruited 389 participants through the Beijing Center of the University of Chicago (68.7% women, 31.1% men, .2% others). Of these, 185 participants were randomly assigned to the spoken condition and 204 to the written condition.

Material and Procedure

The study was conducted in the lab and the procedure was the same as in Study 2. Participants received eight reasoning problems in Chinese, four with a conflict between a relatively more intuitive and a relatively more analytic solution, and four control problems without such conflict (e.g., "How many cubic feet of sand are there in a sandbox that is 1′ deep 1′ wide × 1′ long?"). The presentation order of problems was randomized across participants. For the full set of problems see Supplemental Materials Method Table S3. In the spoken modality, the problems were read aloud in a neutral tone by one of two male speakers both native Chinese speakers from Beijing. We used a 2 (Modality: written, spoken) × 2 (Item type: conflict, no-conflict) mixed-factorial design, with modality as the between-subjects factor and item type as the within-subject factor.

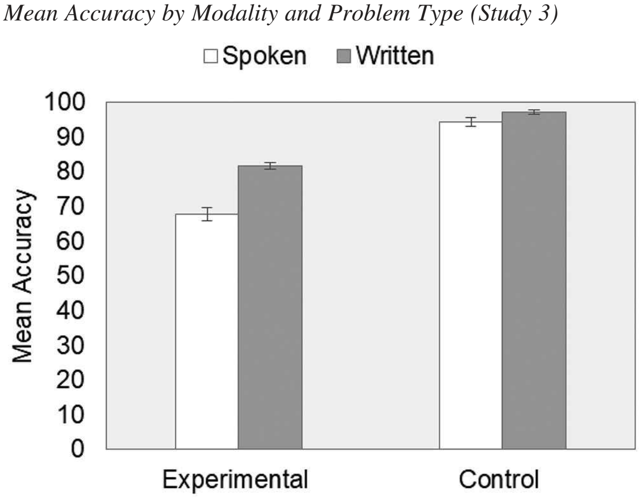
Results and Discussion

In contrast to Studies 1 and 2 where experimental and control items were analyzed separately, here we analyzed them jointly because they were equal in number. We expected that the written modality would lead to more analytically correct responses for the conflict problems than the spoken modality, but have less of an influence on the control problems. For the dependent variable, we focused on the difference in performance between the conflict and control problems. Our theory predicts a bigger difference in the spoken than the written modality condition.

As predicted, the difference was bigger in the spoken than in the written condition (see Figure 3). In the spoken modality, accuracy was 26 percentage points lower for conflict compared with no-conflict problems. In contrast, in the written modality the corresponding difference was only 15 percentage points. This interaction between problem type and language modality was significant, $F(1, 387) = 20.87$, $p < .001$, $\eta_p^2 = .05$.

Pairwise comparisons showed a significant difference between conflict and no-conflict problems for the written condition, $F(1, 387) = 87.17$, $p < .001$, $\eta_p^2 = .18$, as well as for the spoken condition, $F(1, 387) = 231.03$, $p < .001$, $\eta_p^2 = .37$. There was also a main effect of modality, $F(1, 387) = 47.47$, $p < .001$, $\eta_p^2 = .11$, and a main effect of problem type, $F(1, 387) = 304.36$, $p < .001$, $\eta_p^2 = .44$.

Figure 3
Mean Accuracy by Modality and Problem Type (Study 3)



Note. Error bars illustrate standard errors of the mean.

In the data we report above, 12% of participants had one or more missing values. This was due to technical issues that some participants in the spoken condition experienced. They reported that they could not hear the recording occasionally due to Internet connection problems. To make sure that the results are not due to this issue, we conducted additional tests where we restricted the analysis to participants who had a complete data set.

The pattern of the results with this restricted data set mirrors the pattern of the results with the full data set. Accuracy in the spoken condition was 25 percentage points lower for conflict compared with no-conflict problems, while in the written modality the corresponding difference was 16 percentage points. This interaction between problem type and language modality was significant, $F(1, 340) = 17.65$, $p < .001$, $\eta_p^2 = .05$.

Pairwise comparisons showed a significant difference between conflict and no-conflict problems for the written condition, $F(1, 340) = 105.97$, $p < .001$, $\eta_p^2 = .24$, as well as the spoken condition, $F(1, 340) = 193.97$, $p < .001$, $\eta_p^2 = .36$. There was also a main effect of modality, $F(1, 340) = 30.78$, $p < .001$, $\eta_p^2 = .08$, and a main effect of problem type, $F(1, 340) = 298.58$, $p < .001$, $\eta_p^2 = .47$.

In summary, the results of Study 3 generalize the effect of modality on thinking performance to a Chinese population, using a language with a writing system that is different from English. It also provides a direct comparison of conflict and control problems.

Study 4

Studies 1 to 3 demonstrate higher response accuracy in solving verbal conflict problems that are communicated in written rather than spoken form. Therefore, the difference in performance supports the idea that the spoken modality makes more accessible heuristic intuition than the written modality. However, there is an alternative explanation to these findings, which suggests that they reflect a simple presentation format effect. Spoken language tends to be transient while written language can be reviewed and reanalyzed. Such presentation format might provide an advantage to the written condition. Indeed, studies have shown that a spoken presentation format compared with a written presentation format can hurt thinking performance as it taxes working memory by increasing difficulty in keeping the problem information in mind (see Gilhooly et al., 2002). Such differences in presentation format could cause our effect, but it would not reflect a modality effect. It would reflect a presentation format effect. If this is the reason for the findings of Studies 1 to 3, then they do not support our account that modality differentially affects thought.

However, there are three main reasons to doubt this alternative account. First, the one-sentence trivia questions used in Study 1 were relatively simple, so it is unlikely that the modality effect was due to additional working memory load in the spoken modality to store and process the information. Second, in Study 1, where modality was manipulated within participants and order was counterbalanced, we observed a modality by order interaction whereby the modality effect was present when the spoken problems were presented first but absent when the written problems appeared first.

This result is consistent with our prediction that the written modality engages an analytic mindset, which could have spilled over when solving the spoken problems. If the modality effect was due to the additional memory load associated with the spoken modality, then the performance for the spoken problems should have been worse than for the written problems irrespective of presentation order.[1] Third, in all our studies, performance in the control problems was comparable across modality conditions. If the modality effect was merely due to additional difficulty in processing and using the problem information it should have also been manifested in the control problems.

Study 4 was designed to test the modality dependence theory under conditions that rule out the alternative explanation that is based on presentation format. To do that we presented the written problems in a sequential way thereby mimicking the sequential manner of the spoken problems. If the findings in Studies 1 to 3 are merely due to the presentation format, then the effect of modality should disappear in Study 4. If, however, the effect of modality persists, then it would support the account that the two modalities differentially influence thinking performance.

Method

Power Analysis

We estimated the number of participants required by conducting an a priori power analysis for a one-way ANOVA using *G*Power*. We used the following estimates: $\alpha = .05$, power = .95, $f = .20$ (based on Study 2), number of groups = 2. This analysis revealed a minimum sample of 328 participants. We recruited more participants to prevent a reduction in statistical power due to possible exclusions.

Participants

Participants were native English-speaking U.S. residents and were randomly assigned to the spoken or written condition. We requested 410 participants through Prolific (www.prolific.ac). Study 4 was conducted online, so we included three attention checks. We excluded anyone who failed one or more attention checks (24 participants, 5.9%). The results presented below are based on the remaining 386 participants (176 women, 207 men, 3 unknown, $M_{\text{age}} = 40.2$ years, age range = 21 to 80 years). Of these, 116 were in the spoken condition, 123 in the written sequential condition, and 147 in the standard written condition.

Material and Procedure

Study 4 used the same materials and procedure as Study 2. In addition to the spoken and written modality conditions, we included a written-sequential condition that simulated the transient presentation of spoken language. To do this, we presented the written problems in separate phrases on consecutive screens so that participants could not reread them. Participants had to proceed to the next part of the problem by clicking on a button labeled next. Once they pressed the next button they were not allowed to go back. In the spoken modality, we presented the audio information, and just like in the written modality, participants controlled the next button to proceed to the response page. Crucially, participants could not anticipate the question before the final part disappeared. We decided to use this text segmentation reading method for two reasons. First, this method represents the most natural and ecologically valid way of equating the spoken and written language modalities and is widely used in digital media (Szarkowska & Gerber-Morón, 2018). Second, research suggests that reading, like listening, is largely serial and incremental (see Rayner & Clifton, 2009).

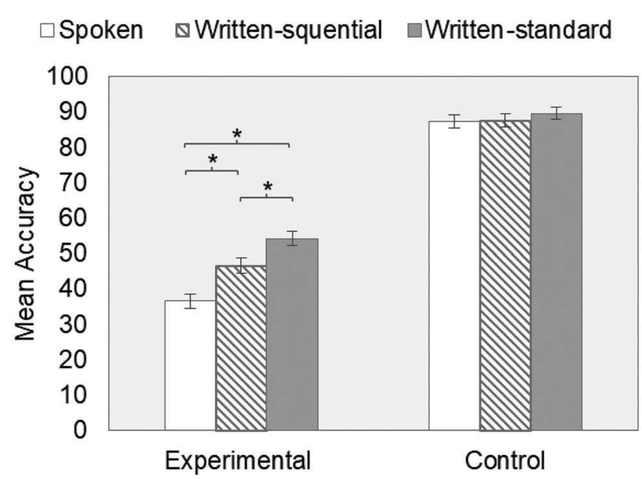
In each modality condition, we presented participants with five conflict reasoning problems and three control problems (see Supplemental Materials Method Table S2). In the spoken modality, the problems were read aloud by two male native English speakers with a standard American English accent. Participants then filled the Need for Cognition and Faith in Intuition scales as in Study 1 (see Supplemental Materials Results Table S2). Lastly, participants indicated their age, gender, education level, and employment status (see Supplemental Materials Results Table S7).

Results and Discussion

Conflict Problems

Participants in the written-sequential condition and the standard written condition were significantly more likely to solve the problems correctly ($M_{\text{Written-sequential}} = .47$, 95% CI [.42, .51], SD = .24; $M_{\text{Written-standard}} = .54$, 95% CI [.50, .58], $SD = .25$) than were participants in the spoken condition ($M_{\text{Spoken}} = .37$, 95% CI [.32, .41], $SD = .23$), Welch's $F(2, 251.39) = 17.40$, $p < .001$, $d_{\text{Cohen}} = .60$; $H(2) = 31.92$, $p < .001$, $d_{\text{Cohen}} = .58$ (see Figure 4). Crucially, pairwise tests, adjusted for multiple comparisons, revealed a significant difference between the written-sequential and spoken conditions ($M_{\text{Diff}} = -.10$, 95% CI [−.03, −.18], $SE = .03$, $p = .004$, $d_{\text{Cohen}} = .43$). There was also a significant difference between the written-standard and spoken conditions ($M_{\text{Diff}} = -.18$, 95% CI [−.25, −.11], $SE = .03$, $p < .001$, $d_{\text{Cohen}} = .73$) and a smaller significant difference between the written-sequential and the written-standard conditions ($M_{\text{Diff}} = .08$, 95% CI [.15, .01], $SE = .03$, p = .034, $d_{\text{Cohen}} = .31$).

Figure 4
Mean Accuracy by Language Modality and Problem Type (Study 4)



Note. Error bars illustrate standard errors of the mean.
* $p < .05$.

Control Problems

Again there was no significant difference in accuracy between the modalities for the control problems ($M_{\text{Written-sequential}}$ = .88, SD = .20, 95% CI [.84, .91], $M_{\text{Written-standard}}$ = .90, SD = .17, 95% CI [.86, .93], M_{Spoken} = .87, SD = .21, 95% CI [.84, .91]), Welch's $F(2, 249.80) = .49$, $p = .616$, $d_{\text{Cohen}} = .11$; $H(2) = 2.11$, $p = .348$, $d_{\text{Cohen}} = .03$. This shows that the spoken modality did not induce an overall deterioration in performance, and that the reduced accuracy in the spoken modality was unique to problems that involved a conflict between heuristic intuitive and analytic thought.

In summary, these results replicate the findings of Studies 1 to 3 and extend the modality effect to a written presentation format that is sequential. Therefore, participants in both modality conditions had to keep the information in mind and were unable to revisit it. This finding, then, supports the idea that language modality influences thinking performance and speaks against the possibility that this effect is simply explained by increased difficulty in processing the problem information in the spoken condition due to its ephemeral nature.

Study 5

Study 5 expanded the scope of the investigation by examining deductive reasoning (Henle, 1962). A central aspect of logical reasoning is validity, which is strictly a formal property that is independent of meaning and truth. A logical argument is valid if and only if its conclusion follows from the premises. For example, the following syllogism is logically invalid:

Premise 1. All living things need water
Premise 2. Roses need water
Conclusion. Therefore, roses are living things

Even though the premises and the conclusion are true, this syllogism is invalid because the conclusion does not follow from the premises. However, because people believe the conclusion, they tend to judge the syllogism as logically valid. More generally, truth-value negatively impacts validity judgments when validity and truth conflict (Evans et al., 1983). People are more likely to judge valid arguments as invalid when the conclusion is unbelievable, and they are more likely to judge invalid arguments as valid when the conclusion is believable. This is called "belief bias" (Wilkins, 1929).

Study 5 capitalized on this bias which appears to be driven by heuristic intuition. The perception of truth is claimed to be an intuitive process that connects to our beliefs (Gilbert, 1991). To reduce the belief bias when judging validity, most people must recruit analytic processes and focus on formal, analytic aspects of the syllogism (Evans & Curtis-Holmes, 2005), while only few people with high cognitive ability can automatically apply logical principles (Raoelison et al., 2020). Therefore, if thinking performance is modality dependent, and listening cues relatively more heuristic intuition, then the belief bias should be greater in the spoken than in the written modality.

Method

The prediction, study design, sample size, and analyses were preregistered on AsPredicted.org, see aspredicted.org/z78av.pdf.

Power Analysis

We conducted an a priori power analysis for an independent samples t test (two-tailed) using the following estimates: $\alpha = .05$, power = .80, $d_{\text{Cohen}} = .50$ (based on piloting), $df = 1$. This analysis revealed a minimum sample of 128 participants. We recruited more participants to prevent reduction in statistical power due to possible exclusions.

Participants

Participants were native English-speaking U.S. residents and were randomly assigned to the spoken or written condition. We requested 140 participants through Prolific (www.prolific.ac) and collected data from 156 who passed the screening tests, in anticipation of some attrition. Seven participants (4.5%) were excluded because they failed one or more attention checks. We analyzed the data of the remaining 149 participants (83 women, 62 men, 4 other, M_{age} = 31.4 years, age range = 18–60 years), where 80 were in the spoken condition and 69 were in the written condition.

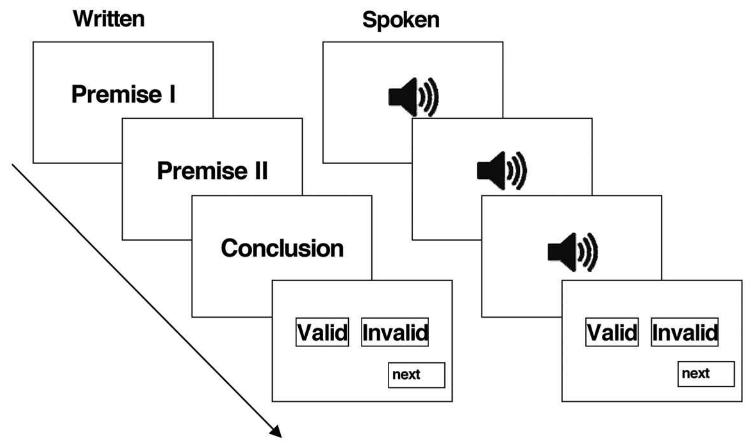
Material and Procedure

Participants received 12 syllogisms, each comprising two premises (four or five words) and a conclusion (five or six words). Six syllogisms were valid but had an unbelievable conclusion, and six were invalid but had a believable conclusion. We followed West et al. (2008) in this design, in which the believability of the information was inconsistent with the logical format of the syllogisms in both types. To solve such problems correctly, participants need to put aside their knowledge of facts and reason based solely on the relationship between the premises and the conclusion. Supplemental Materials Method Table S4 presents the full set of items. Before beginning the task, participants received an explanation of logical validity (see Supplemental Materials Method for the wording of the instructions). To ensure participants' understanding of the task, following the instructions participants were asked two task comprehension questions. If participants incorrectly responded to one of these questions, they were redirected to the instruction page. Then participants practiced two examples of each syllogism type followed by feedback on their performance (see Supplemental Materials Method for the wording of the feedback).

Participants in the spoken condition heard the syllogisms spoken in a neutral tone by two male native English speakers from the Chicago area with a standard American English accent. Participants in the written sequential and spoken conditions received the syllogisms in a transient form: first premise (four words), second premise (four to five words), and conclusion (five to six words), each presented on a separate screen, and they were not allowed to go back (see Figure 5). Following each written and spoken information, participants had to proceed to the next part of the syllogism by clicking on a button labeled "next".

The written condition is analogous to the sequential written condition tested in Study 4 and hence an eventual modality effect cannot be ascribed to differences in presentation format as in both the spoken and the written condition participants have to keep the information in mind and integrate it to evaluate the validity of the conclusion. Participants' task was to judge the syllogism's validity by

Figure 5
Illustration of the Procedure Used in Study 5



selecting either "*Valid"* or "*Invalid*." The presentation order of the syllogisms was randomized across participants.

Previous research has found performance differences with syllogistic reasoning between spoken and written presentation formats, but not in the context of the belief bias (Gilhooly et al., 1993). Participants provided more correct responses in the written condition. However, these findings could be fully explained as a presentation format effect, not necessarily a modality effect. This is because the syllogisms in the written condition were not presented sequentially but were presented in their entirety without time limits. This gave a clear advantage to participants in the written condition over the spoken condition as they did not have to keep the written premises and conclusions in working memory while evaluating validity. In our case, participants received the premises and the conclusion sequentially and did not have access to them when they evaluated validity. Given that in both modality conditions participants had to keep this information in mind, our procedure provides an opportunity to evaluate a modality effect as distinct from a presentation effect.

After completing the syllogism task, participants answered the following questions: "How many problems out of 12 do you think you solved correctly?" (0 to 12 scale), and "Overall, how difficult was it for you to solve the problems?" (slider scale: 1 = *not at all difficult*, to 101 = *extremely difficult*). Then, they rated their agreement with the following statements: "While reading the problems, I was fully absorbed," "I enjoyed reading the problems," "I felt totally involved in reading the problems," "While reading the problems, I had the impression that time was passing quickly," and "I found the task extremely rewarding" (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). In the spoken condition, the text had "listening to" instead of "reading" (see Supplemental Materials Results Table S1 for descriptive statistics).

For exploratory purposes, participants then received three CRT type problems in written form. The goal was to explore whether solving syllogisms in the two modalities prompts different levels of cognitive fatigue (Ackerman & Kanfer, 2009). Previous evidence demonstrates that increased cognitive fatigue deteriorates performance in subsequent reasoning tasks (e.g., Inzlicht & Schmeichel, 2012; Mani et al., 2013; Timmons & Byrne, 2019). Therefore, if modality differentially taxes cognitive resources and consequently influences cognitive fatigue, then this might influence performance on the subsequent CRT task.

Following the CRT task, participants filled out the Need for Cognition and Faith in Intuition scales (see Supplemental Materials Results Table S2 for descriptive statistics). Finally, participants answered questions concerning their prior experience with solving syllogisms, formal education in logic, their age, gender, and educational level (see Supplemental Materials Results Table S7 for descriptive statistics; see Supplemental Materials Results for exploratory analyses).

Results and Discussion

Experience With Solving Syllogisms

First, we evaluated whether participants in the two modality conditions differed in their past experience in solving syllogisms or their formal education in logic. We found no differences for experiences, $\chi^2(1, N = 149) = .02$, $p = .886$, $\phi = .01$, or formal education in logic, $\chi^2(1, N = 149) = .14$, $p = .709$, $\phi = .03$.

Validity Judgments

We computed an accuracy index for each participant by averaging the rate of correct acceptance of valid problems and correct rejection of invalid problems across the 12 syllogisms. Preliminary analyses revealed that speaker identity did not affect accuracy ($F < 1$); therefore, we dropped this factor from the analyses.

As predicted, participants in the written condition were significantly more accurate in judging validity than participants in the spoken condition ($M_{\text{Written}} = .60$, 95% CI [.55, .65], $SD = .20$, $M_{\text{Spoken}} = .54$, 95% CI [.50, .58], $SD = .18$), Welch's $F(1, 136.39) = 4.26$, $p = .041$, $d_{\text{Cohen}} = .34$, $U = 3{,}267.50$, $z = 1.95$, $p = .051$, $d_{\text{Cohen}} = .18$.

This shows that participants in the spoken condition were more affected by the belief bias, suggesting that the spoken modality engages comparatively more heuristic intuition than the written modality.

Performance in the CRT Task

The CRT problems followed the syllogism task and were received by all participants in written form. This task was included for exploratory purposes. Its aim was to assess whether the modality in which the syllogisms were presented differentially taxed cognitive resources. If it did, then performance on the subsequent CRT problems could be affected as studies suggest that cognitive fatigue deteriorates reasoning performance. There was no difference in CRT performance across the modalities (M_{Written} = .68, 95% CI [.59, .77], SD = .38, M_{Spoken} = .63, 95% CI [.54, .71], SD = .40), Welch's t(145.72) = .88, p = .378, d_{Cohen} = .15, U = 2,971.50, z = .87, p = .387, d_{Cohen} = .20. This suggests that the modality effect on syllogistic reasoning is not due to increased cognitive fatigue in the spoken modality. Therefore, the way we presented the spoken and written information was successful in equating the conditions in terms of working memory load.

CRT performance correlated with performance in the syllogism task, r(147) = .285, $p < .001$; however, the magnitude of this correlation did not differ across conditions (Written: r(78) = .335, p = .002, Spoken: r(104) = .309, p = .001; z = .19, p = .848, q_{Cohen} = .03). This suggests that the two tasks revealed consistent individual differences in reasoning: people that perform better with syllogisms also perform better on the CRT (e.g., Toplak et al., 2014).

Because the written condition involved a sequential presentation, the modality effect cannot be ascribed to a simple presentation advantage in the written condition. Indeed, the results from the CRT task provide converging indirect evidence, showing that cognitive fatigue after the syllogism task was not affected by the modality of performing it.

In summary, Study 5 demonstrated that judgments of logical validity are affected by modality. Participants were more susceptible to the belief bias when they heard the syllogisms than when they read them. To the extent that the belief bias is linked to intuition (Evans & Curtis-Holmes, 2005), the present results suggest that thinking from spoken premises relies relatively more on heuristic intuition than thinking from written premises.

General Discussion

Our results suggest that thinking performance is modality dependent. When people answered spoken trivia questions, they were less likely to detect semantic distortions and more likely to fall for the Moses illusion than when they answered the same trivia questions in written form. When people solved spoken riddles, they responded relatively more intuitively than when they solved the same riddles in written form. When judging logical validity, people were more affected by the belief bias in the spoken modality, suggesting that heuristic intuition plays a larger role in that modality than in the written one. Clearly, thinking is not modality independent.

We found evidence for a modality effect on thinking performance by testing two very different languages namely English and Chinese, which speaks to the generalizability of the effect across languages and populations. Furthermore, we tested problems involving different levels of complexity: simple detection of semantic anomalies, relatively more complex verbal riddles, and harder multipremised syllogisms. The fact that the modality effect was present using simple one-sentence semantic anomalies speaks against the idea that the effect is due to an increased difficulty of processing the information in the spoken modality. Furthermore, the effect persisted when presenting written problems in a sequential form mimicking the presentation format of spoken problems. This further helps to rule out a simple explanation that the effect is due to differences in presentation format. Our results, therefore, suggest that the very comprehension of spoken problems affects thinking performance differently than the comprehension of written problems, because the spoken modality privileges heuristic intuition, while the written modality privileges analytic thought.

Although problems that require logical reasoning tend to benefit from analytic processing, correct responding can also be generated spontaneously without deliberation. For example, one can solve reasoning problems by automatically applying logical or mathematical principles that have been internalized (De Neys, 2006). Such logical intuition should apply predominantly to problems whose solutions involve the application of logico-mathematical principles (Sinayev & Peters, 2015). With the exception of Study 5, the present studies tested verbal problems that do not require the application of logico-mathematical principles (Sirota et al., 2021). Therefore, performance in Studies 1 to 4 is unlikely to be relevant to logical intuition.

Alternative Explanation

An alternative account for the impact of modality on thinking might be that listening is more cognitively demanding than reading, hence increasing the load on working memory (Klingner et al., 2011). One could imagine two versions of this account. One version is that because spoken language is transient, people rely more on working memory to keep track of the information thereby depleting working memory resources (Gilhooly et al., 2002). This could reduce the resources available to think deliberatively. Our results speak against this account because they demonstrate that the modality effect persists even when transience of the language is controlled.

A second version could suggest that cognitive demand is higher in the spoken modality for reasons other than its transient nature. The literature does not provide evidence for this account. If anything, research on short term memory for spoken and written language suggests the opposite. Studies using relatively simple sequential working memory tasks, such as digit span tasks, found that performance for spoken stimuli tends to be better than for written stimuli (e.g., Greene, 1992; Penney, 1989). Similarly, research testing modality effects with a working memory task in which sequences of letters are presented shows better performance accuracy with spoken than with written language (Amon & Bertenthal, 2018). Had spoken language been more cognitively demanding, performance in these working memory tasks should have been worse with spoken language, not better.

Theoretical Contribution

We propose that thinking performance based on spoken problems is systematically different than thinking performance based on written ones because the spoken modality privileges heuristic intuitive thinking, while the written modality privileges analytic thinking. This is motivated by the modalities' distinct engagement of controlled processing, developmental trajectory, and the context in which they are likely to be used over time. Written language is assumed to engage more controlled processing than spoken language (Varao Sousa et al., 2013). Furthermore, spoken language is acquired spontaneously and intuitively while written language is acquired via formal instruction (Liberman, 1992; Liberman & Whalen, 2000; Pinker, 1994). Finally, spoken language is typically used in informal conversations while written language is often used in formal contexts. Therefore, it is likely that the modality effect is related to differences in the underlying cognitive processing, developmental trajectory, and context of use of these modalities.

Our finding contributes to the theoretical understanding of the relationship between language and thought broadly defined. Research has identified mild effects on cognition of structural elements of language, such as grammatical gender, as well as other aspects of language including metaphors about time or motion (Boroditsky, 2001; Gumperz & Levinson, 1991; Sapir, 1929; Sera et al., 1994; Whorf, 1956; but see Gleitman & Papafragou, 2012). Our theory and the supporting studies add an important novel dimension to our understanding of how language might influence cognition: through its modality.

Our research also contributes to theories of thinking, which tend to be concerned with the conditions under which thought involves deliberation and intuition (Evans & Curtis-Holmes, 2005; Sloman, 2002; Thompson, 2012). Research has demonstrated that individual differences as well as contextual factors can impact the extent to which thinking involves deliberation and intuition (De Neys, 2006; Sorrentino & Stanovich, 2002; Toplak et al., 2014). For example, high cognitive ability and instructions to focus on the logical nature of the task can promote correct performance (Evans et al., 1994; Newstead et al., 1992; Sorrentino & Stanovich, 2002) while time pressure favors heuristic intuition (Evans & Curtis-Holmes, 2005; Finucane et al., 2000). Our finding suggests that language modality might modulate such thinking processes, leading spoken language to privilege heuristic intuitive thinking compared with written language.

Furthermore, the vast majority of thinking research, as well as research on judgment and decision making, presents instructions and linguistic stimuli in the written modality under the implicit assumption that modality is immaterial for evaluating theories of thinking and decision making. Our results illustrate that this methodological practice might lead to a systematic underestimation of the extent to which people respond intuitively to problems. The magnitude of the misestimation could be significant because everyday thinking is often based on spoken communications.

Implications

Our findings have potentially important implications in a variety of domains. Because they challenge the fundamental assumption that thought is modality independent, our findings could impact the way research is conducted. Research in psychology, sociology, political science, economics, and other social sciences often involves instructions and stimuli. However, most choose the modality of the language by convenience, precisely because of the implicit assumption that modality is immaterial for most tasks. For example, some surveys present questions in writing while in others the questions are spoken directly to participants. The modality effect suggests that providing surveys in the spoken modality responses might be relatively more intuitive. Thus, a public opinion survey about illegal immigration might tap into feelings when conducted orally, while a written format might involve fewer such emotional considerations. Therefore, such disciplines might rely on our findings for a more reasoned selection of the modality to not bias results.

Our findings also carry potential implications for any domain where thinking and reasoning is central such as medicine, business, and the law. For example, legal reasoning is crucial for the practice of law, which applies both rules of deductive reasoning and analogical reasoning via cases (Pashler & Ellsworth, 2013). However, it does not consider language modality as a factor. A legal brief makes the same argument whether it is read or heard, but it might not have the same impact. Our discovery suggests that reading an argument would lead to more analytic outcomes, whereas hearing it would give more consideration to heuristic intuition.

Conclusion

In the history of humanity, written language developed thousands of years after spoken language (Houston, 2004). As recently as 200 years ago, only 12% of the world's population was able to read and write (Roser & Ortiz-Ospina, 2018). Our studies demonstrate that the choice between the older and the newer language modality is consequential: Modality directly affects thought, making us more intuitive when we hear a problem while more analytic when we read it. This finding should inform not only theories of thinking, but also practices in which the nature of thought is consequential such as policy making, judicial reasoning, and medical decision making.

References

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163–181. https://doi.org/10.1037/a0015719

Amon, M. J., & Bertenthal, B. I. (2018). Auditory versus visual stimulus effects on cognitive performance during the N-back task. *Cognitive Science*. Advance online publication. https://cogsci.mindmodeling.org/2018/papers/0255/index.html

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Arieli, S., & Sagiv, L. (2018). Culture and problem-solving: Congruency between the cultural mindset of individualism versus collectivism and problem type. *Journal of Experimental Psychology: General*, *147*(6), 789–814. https://doi.org/10.1037/xge0000444

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Braine, M. D. S., O'Brien, D. P., & Braine, M. (Eds.). (1998). *Mental logic*. Psychology Press. https://doi.org/10.4324/9781410603005

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Channon, S., & Baker, J. (1994). Reasoning strategies in depression: Effects of depressed mood on a syllogism task. *Personality and Individual Differences*, *17*(5), 707–711. https://doi.org/10.1016/0191-8869(94)90148-1

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *Journal of Direct Instruction*, *1*(2), 137–149.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724. https://doi.org/10.1037/0003-066X.49.8.709

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*(3), 378–395. https://doi.org/10.3758/BF03193858

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389. https://doi.org/10.1080/13546780542000005

Evans, J. S. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *The European Journal of Cognitive Psychology*, *6*(3), 263–285. https://doi.org/10.1080/09541449408520148

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Ford, M., & Johnson-Laird, P. N. (1985). Mental models: Towards a cognitive science of language, inference, and consciousness. *Language*, *61*(4), 897. https://doi.org/10.2307/414498

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Geipel, J., & Keysar, B. (2021). *Language & thinking* (No. 12761885). Open Science Framework. https://osf.io/wyqh6/?view_only=63c7efa4e0e840e59b3ff9f5f67569dc

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119. https://doi.org/10.1037/0003-066X.46.2.107

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. E. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychological Research & Reviews*, *21*(2), 111–120. https://doi.org/10.1007/s12144-002-1007-x

Gleitman, L., & Papafragou, A. (2012). New perspectives on language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 543–568). Oxford Library of Psychology. https://doi.org/10.1093/oxfordhb/9780199734689.013.0028

Greene, R. L. (1992). *Human memory*. Psychology Press. https://doi.org/10.4324/9781315807195

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. *Current Anthropology*, *32*(5), 613–623. https://doi.org/10.1086/204009

Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own success? *Advances in Cognitive Psychology*, *12*(3), 145–149. https://doi.org/10.5709/acp-0193-5

Hammond, K. R. (1996). *Human judgment and social policy*. Oxford University Press.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of 'motherese'? *Journal of Child Language*, *15*(2), 395–410. https://doi.org/10.1017/S0305000900012411

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, *69*(4), 366–378. https://doi.org/10.1037/h0042043

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), Article 29. https://doi.org/10.1038/466029a

Houston, S. D. (2004). The archaeology of communication technologies. *Annual Review of Anthropology*, *33*(1), 223–250. https://doi.org/10.1146/annurev.anthro.33.070203.143724

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23–30. https://doi.org/10.1037/0096-3445.133.1.23

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*(5), 450–463. https://doi.org/10.1177/1745691612454134

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, *48*(3), 323–332. https://doi.org/10.1111/j.1469-8986.2010.01069.x

Kopp, K., & D'Mello, S. (2016). The impact of modality on mind wandering during comprehension. *Applied Cognitive Psychology*, *30*(1), 29–40. https://doi.org/10.1002/acp.3163

Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel-commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*(3), 242–247. https://doi.org/10.1177/1745691613483477

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0020762

Liberman, A. M. (1989). Reading is hard just because listening is easy. In C. von Euler (Ed.), *Wenner-Gren International Symposium Series: Brain and reading* (pp. 197–205). Macmillan. https://doi.org/10.1007/978-1-349-10732-2_14

Liberman, A. M. (1992). The relation of speech to reading and writing. *Advances in Psychology*, *94*, 167–178. https://doi.org/10.1016/S0166-4115(08)62794-6

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*(6149), 976–980. https://doi.org/10.1126/science.1238041

Margolin, C. M., Griebel, B., & Wolford, G. (1982). Effect of distraction on reading versus listening. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(6), 613–618. https://doi.org/10.1037/0278-7393.8.6.613

Maymin, P. Z., & Langer, E. J. (2021). Cognitive biases and mindfulness. *Humanities & Social Sciences Communications*, *8*(1), 1–11. https://doi.org/10.1057/s41599-021-00712-1

McGregor, W. B. (2009). *Linguistics: An introduction*. Continuum International.

Michael, E. B., Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, *13*(4), 239–252. https://doi.org/10.1002/hbm.1036

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781. https://doi.org/10.1038/nn1077

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284. https://doi.org/10.1016/0010-0277(92)90019-E

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. https://doi.org/10.1037/0033-295X.108.2.291

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 476–492. https://doi.org/10.1037/0278-7393.22.2.476

Oldrati, V., Patricelli, J., Colombo, B., & Antonietti, A. (2016). The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia*, *91*, 499–508. https://doi.org/10.1016/j.neuropsychologia.2016.09.010

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988–1010. https://doi.org/10.3758/BF03196730

Pashler, H., & Ellsworth, P. C. (2013). Legal reasoning, psychological perspectives. In H. Pashle (Ed.), *Encyclopedia of the mind* (pp. 469–470). SAGE Publications, Inc. https://doi.org/10.4135/9781452257044.n175

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398–422. https://doi.org/10.3758/BF03202613

Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: Evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*, *62*(4), 380–391. https://doi.org/10.1016/j.jml.2010.02.005

Pinker, S. (1994). *The language instinct: How the mind creates language*. Penguin Books. https://doi.org/10.1037/e412952005-009

Raoelison, M., Thompson, V. A., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition*, *204*, Article 104381. https://doi.org/10.1016/j.cognition.2020.104381

Rayner, K., & Clifton, C., Jr. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, *80*(1), 4–9. https://doi.org/10.1016/j.biopsycho.2008.05.002

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31–74. https://doi.org/10.1111/1529-1006.00004

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading* (2nd ed.). Psychology Press. https://doi.org/10.4324/9780203155158

Regev, M., Honey, C. J., Simony, E., & Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *The Journal of Neuroscience*, *33*(40), 15978–15988. https://doi.org/10.1523/JNEUROSCI.1580-13.2013

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*(1), 38–71. https://doi.org/10.1037/0033-295X.90.1.38

Rips, L. J. (1994). The psychology of proof: Deductive reasoning in human thinking. *Choice Reviews Online*, *32*(02), 32–1231. https://doi.org/10.5860/CHOIC

Roser, M., & Ortiz-Ospina, E. (2018). *Literacy*. Published online at OurWorldInData.org. https://ourworldindata.org/literacy

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386. https://doi.org/10.1016/S1364-6613(02)01958-7

Sapir, E. (1929). The status of linguistics as a science. *Language*, *5*(4), Article 207. https://doi.org/10.2307/409588

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Sera, M. D., Berge, C. A. H., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*(3), 261–292. https://doi.org/10.1016/0885-2014(94)90007-8

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *6*, Article 532. https://doi.org/10.3389/fpsyg.2015.00532

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., & Marshall, A. C. (2021). Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making*, *34*(3), 322–343. https://doi.org/10.1002/bdm.2213

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 379–396). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.024

Sorrentino, R. M., & Stanovich, K. E. (2002). Who is rational and when? Who is rational?: Individual differences in reasoning. *The American Journal of Psychology*, *115*(1), 132–141. https://doi.org/10.2307/1423680

Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS ONE*, *13*(6), Article e0199331. https://doi.org/10.1371/journal.pone.0199331

Thompson, V. A. (2012). Dual-process theories: A metacognitive perspective. In T. Minds (Ed.), *Dual processes and beyond* (pp. 171–196). Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199230167.003.0008

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Timmons, S., & Byrne, R. M. J. (2019). Moral fatigue: The effects of cognitive fatigue on moral reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *72*(4), 943–954. https://doi.org/10.1177/1747021818772045

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, *64*(1), 231–255. https://doi.org/10.1146/annurev-psych-113011-143812

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, 109–118. https://doi.org/10.1016/j.cognition.2016.01.015

Usher, M., Russo, Z., Weyers, M., Brauner, R., & Zakay, D. (2011). The impact of the mode of thought in complex decisions: Intuitive decisions are better. *Frontiers in Psychology*, *2*, Article 37. https://doi.org/10.3389/fpsyg.2011.00037

Varao Sousa, T. L., Carriere, J. S. A., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in Psychology*, *4*, Article 892. https://doi.org/10.3389/fpsyg.2013.00892

Varnum, M. E. W., Grossmann, I., Kitayama, S., & Nisbett, R. E. (2010). The origin of cultural differences in cognition: The social orientation hypothesis. *Current Directions in Psychological Science*, *19*(1), 9–13. https://doi.org/10.1177/0963721409359301

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*(4), 930–941. https://doi.org/10.1037/a0012842

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives de Psychologie*, *102*, Article 83. http://psycnet.apa.org/psycinfo/1929-04403-001

Received December 8, 2021
Revision received June 29, 2022
Accepted September 20, 2022 ■