Check for updates

The effect of writing script on efficiency and metacognitive monitoring in inferential word learning

Leona Polyanskaya[1,2,3] · Dina Abdel Salam El-Dakhs[3] · Ming Tao[4] · Fengfeng Chu[4] · Mikhail Ordin[2]

Abstract

The writing system – the transparency of orthography in alphabet-based systems and differences between logographic and phonetic-based systems – can affect the efficiency of inferential word learning when words are introduced visually. It can also shape how people self-evaluate their learning success (we refer to such type of self-evaluation as metacognitive monitoring of word learning). By contrast, differences in metacognition and learning performance do not emerge when words are presented auditorily. To measure metacognition, we assessed retrospective confidence by asking participants to rate their certainty about the correctness of their responses. As this direct question raises a person's conscious awareness of how well they have learned a particular lexical unit, it allowed us to measure those aspects of metacognition that are modulated by consciousness. Such consciousness comes into play when a word is associated with an object. Differences in conscious awareness of the word learning success when words are represented visually make differential demands on word learning across languages and modalities. The observed differences between populations using different writing systems and between perceptual modalities may potentially modulate the effectiveness of vocabulary acquisition activities during foreign language learning.

Keywords Metacognition · Writing script · Writing system · Literacy · Vocabulary · Word learning

✉ Mikhail Ordin
mikhail.ordin@icnas.uc.pt

1 Department of Psychology and Educational Sciences, Universidade de Coimbra, Coimbra, Portugal

2 Laboratory of Language, Metacognition and Decision-making, Coimbra Institute for Biomedical Imaging and Translational Research, Centre for Nuclear Sciences Applied to Health, Universidade de Coimbra, Coimbra, Portugal

3 Applied Linguistics Research Lab, College of Humanities and Sciences, Prince Sultan University, Riyadh, Saudi Arabia

4 The Second School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China

﹛ Springer

Introduction

Metacognitive monitoring – the ability to track and evaluate one's own cognitive performance – in a particular domain can be modulated by individual experience within that domain (Carpenter et al., 2019; Rademaker & Pearson, 2012; Song et al., 2011). This is evident even in some realms of high-level cognition, such as language processing, which relies on a complex interplay of multiple neuro-cognitive mechanisms. An individual's depth of experience in linguistic processing influences their meta-monitoring of these same cognitive mechanisms, which are also engaged during the processing of linguistic information. Compared to monolinguals, for instance, bilinguals show enhanced metacognition in statistical learning tasks (e.g., those using a Saffran-style paradigm) (Ordin et al., 2020). Moreover, bilinguals whose languages are typologically different may exhibit a further metacognitive advantage – one that surpasses the effects of bilingualism in and of itself – over bilinguals with typologically similar languages. This hypothesis was tested by Polyanskaya et al. (2022), who evaluated metacognitive monitoring in statistical learning tasks on language-like material in the visual and auditory perceptual modalities in three linguistic populations: bilinguals with syntactically different languages (Basque and Spanish); bilinguals with syntactically similar languages (Catalan and Spanish), and Castilian Spanish monolinguals. The hypothesis was fully confirmed *in the auditory modality*, with Basque bilinguals demonstrating the highest metacognitive ability and monolinguals the lowest. Polyanskaya et al. (2022) suggested that the Basque and Spanish languages required distinct cognitive processing strategies due to their typological differences, but that both strategies require monitoring and cognitive performance evaluation, thereby naturally training the user's metacognitive abilities. Syntactically similar languages – Catalan and Spanish – can be parsed by the same cognitive processing strategy, and metacognition of Catalan speakers is not further modulated by typological distance, with the effect of bilingualism in and of itself still leading to better metacognitive monitoring compared to monolinguals. *In the visual modality*, however, the hypothesis was not confirmed. No difference in metacognition was observed between monolingual Spanish speakers and Basque bilinguals, and metacognition in each population was substantially higher than that in the Catalan bilingual population. These findings are shown in Fig. 1.

Polyanskaya et al. (2022) proposed a post-hoc explanation for the patterns observed in the visual modality, which is based on the assumption that statistical learning in different perceptual modalities can be served by differential sets of neuro-cognitive mechanisms (Conway & Christiansen, 2005, 2006; Frost et al., 2015; Polyanskaya, 2022; Thiessen et al., 2013). Hence, metacognitive processes must track and evaluate the effectiveness of these varying cognitive mechanisms across modalities. The processing of visual linguistic information relies on reading skills, which are known to be modulated, among other things, by spelling conventions (Bolger et al., 2005; Ferrand, 2007; Seymour et al., 2003; Thorstad, 1991). Orthography in Spanish and Basque is transparent, with direct sound-to-letter mapping bi-directionally (one can spell an unknown word when it is heard, and there is an unambiguous way of producing a word that is represented in letters). Orthography in Catalan is more opaque, and increased ambiguity in sound-to-letter mapping potentially interferes with metacognitive monitoring (if the rules are less clear, it is more difficult – and less useful – to evaluate how well these rules are observed), discouraging higher metacognitive precision. Hypothetically, the effect of language-specific writing on metacognition in the visual modality may be stronger than or may modulate the effect of typological syntactic properties.
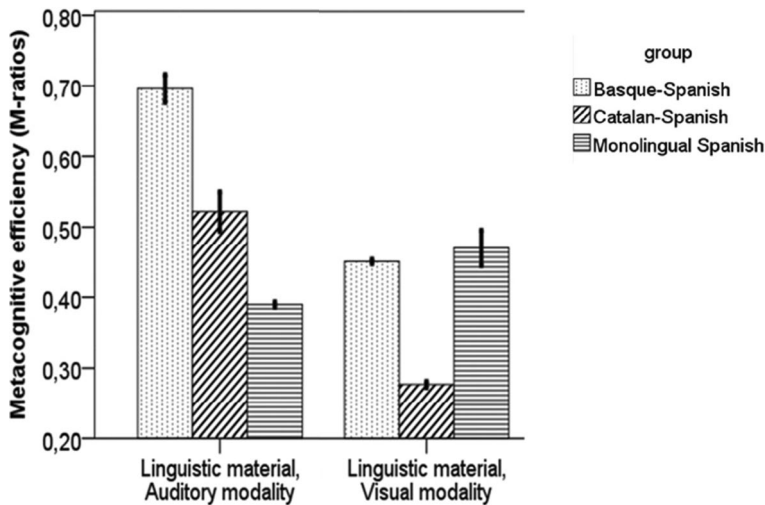
Fig. 1 Metacognition in a statistical learning task in the visual and auditory modalities on linguistic material (figure adopted from Polyanskaya et al., 2022). Metacognition is measures as metacognitive efficiency (see methods for details on the measurement)

The potential effect of writing conventions on metacognition in the visual modality was a post-hoc explanation for an unexpected pattern of results, and it required a hypothesis-driven empirical confirmation. To test this hypothesis experimentally, we decided to focus on the effect of the writing system (logographic in Mandarin; quasi-syllabic abjad in Arabic; alphabetic in Portuguese). Although Arabic is not a purely syllabic language, its writing system (known as *abjad*) represents only consonants, with the vowels inferred based on a variety of diacritics. A typical symbol in Arabic script denotes a consonant followed by a vowel. However, not all vowels are explicitly marked in writing, and diacritics can serve other functions, such as differentiating between word-final and subsequent word-initial consonants represented by the same *abjad* symbol. While the *abjad* intends to capture sounds-to-symbols correspondences, the mapping between pronunciation and symbols is indirect, which makes the orthography more opaque and less systematic than in Portuguese, which has a purely alphabetic writing system with relatively direct sound-to-symbol mapping both for vowels and consonants. We do not claim that Portuguese has the most transparent orthography among alphabetic languages, and we do acknowledge that Spanish, Italian, Basque or Greek orthography, to name just a few, would be more transparent than Portuguese. Relative to Arabic script, however, the orthography in Portuguese is more transparent.

Chinese characters typically include a pair of sub-units called (1) phonetic and (2) semantic radicals, each can vary in complexity. The complexity of phonetic radicals is based on the number of strokes, while the complexity of semantic radical depends on the number of meanings they can convey. Phonetic radicals do not provide systematic cues to pronunciation. The mapping between the symbols and phonemes does not exist, word pronunciations must be memorized, and phonetic radicals serve as memory support for a set of valid variants how the whole word can be produced. The second radical is semantic, and unless the meaning of the character is culturally transmitted from one individual to another, it has to be guessed based on iconicity of the character, pragmatic context, and similarity of the character with

other characters that are already known to the reader. Inferential word learning in logographic systems is akin to learning another communication code, one not closely tied to the spoken language. As inferential word learning in logographic system relies more heavily on memory in addition to phonological awareness, it engaged a wider range of cognitive processes compared to inferential word learning in writing systems that enable symbol-to-sound mapping.

Dong et al. (2021) showed that processing written information in Chinese generates neural activity patterns that differ from those for both Uyghur (Arabic script) and English (alphabetic script). They also observed highly similar neural activation patterns in individuals processing the latter two languages (participants were trilingual Uyghur native speakers with Mandarin and English as second languages). In terms of orthography, English has a closer resemblance to Uyghur than Mandarin, primarily because sound-to-symbol mapping, albeit not transparent, is present in both languages, whereas such mapping in not present in Mandarin. While all three languages have opaque orthography, processing Chinese script probably engages different (and more distributed) neural networks compared to Uyghur abjad and English. This might indicate its reliance on a different and wider range of cognitive mechanisms. Since the output of these mechanisms must be monitored, the additional burden on the metacognitive system offers natural metacognitive training in the visual modality.

We assumed that fundamental differences between writing systems would have a greater influence on metacognition than more subtle differences between opaque and transparent orthography within alphabetic writing system, which is advantageous for *initial* hypothesis testing. If the effect of the script is not observed despite such fundamental differences in writing systems, then further study of more subtle differences in writing on metacognitive monitoring of processing written information would not be warranted.

More specifically, we aimed to test the hypothesis that different writing systems would differentially modulate metacognition in the same language task, depending on whether it involve written information processing. We chose inferential word learning as our task because it occurs in both the visual modality (which involves the processing of written words) and the auditory modality (where such processing is absent). From this hypothesis, we formulated one general and two specific predictions. The general prediction was that metacognition would differ between linguistic populations with different writing systems when the task was performed in the visual modality, but that this difference would not be observed in the auditory modality. The first specific prediction was that the writing system with more transparent sound-to-symbol mapping (Portuguese) would lead to higher metacognition in the task compared to the system with more opaque mapping (Arabic). The second specific prediction was that the writing system which engages more cognitive processes (Chinese), each of which will have to be monitored, would naturally train metacognition in the visual modality leading to enhanced metacognitive skills beyond the effect of sound-to-symbol mapping. Hence, in Chinese population, we expected higher metacognition, due to their logographic writing system, than in Arabic or Portuguese populations. If these predictions are confirmed, we intended to outline pedagogical implications related to explicit language learning and vocabulary building activities.

Method

The study received ethical approval from the Ethics and Deontology Committee of the Faculty of Psychology and Educational Sciences, University of Coimbra (protocol CEDI/FPCEUC:70/0PE). The experiment was presential and informed consent forms were signed and kept on file. The experiments were programmed in PsychoPy. This study was not preregistered.

We recruited 84 Arabic speakers (52 women), 70 Chinese speakers (35 women) and 72 Portuguese speakers (39 women) aged 18 to 30 (mean age: 21.5; median age: 21; there was no age difference between linguistic samples). All participants also had acquired an intermediate knowledge of English through formal education and were university students. Most participants were single, with married participants from Saudi Arabia (2) and Portugal (4). In Portugal and China, the biological sex of participants and research assistants did not always match (in both countries we had male and female assistants involved in the study). In Saudi Arabia, male only male research assistants communicated with males participants, and female research assistants – with female participants.

As a task, we used inferential word learning, defined as the learning of new words using contextual cues, which is a core word learning method in natural language acquisition (Pinker & Jackendoff, 2005). Given that Arabic, Portuguese and Chinese differ both phonotactically and phonologically, a pseudoword that looks and sounds natural in one language might look or sound very weird to speakers of other languages. To remedy this, we created pseudowords tailored to each language, which required that we fine-tune each set of pseudowords to the linguistic characteristics of each specific language. To generate the pseudowords in Portuguese and Saudi Arabic, we used *UniPseudo: A universal pseudoword generator* (New et al., 2024,). For Chinese, we used the pseudoword characters from MELD-SCH (MEgastudy of Lexical Decision in Simplified CHinese), selecting the non-existent characters that yielded the minimum number of false alarms (Tsang et al., 2018). Details on how the nonsense words were constructed, full lists of visual and auditory nonsense words, and all referent images can be found in the supplementary material online.

In the visual modality, participants were presented with two images that differed by a minor detail. For instance, they might see a table and a table with a basket on it, and read the sentence on the screen instructing them to select a table with a "dax" (*dax* being one of the pseudowords). In semantically ambiguous situations, people assume that the speaker chooses the words to be informative (Frank & Goodman, 2014; Quine, 1960); this assumption helps them to resolve the ambiguity and – in the case just cited – to pick the table with a basket. Concurrently, this experience also teaches them the referent for a new pseudoword. In one of the subsequent trials, participants might be presented with the same table with a basket and, in a separate image, the table with a lamp on it, and they were instructed to find a table with a "tife." As participants already knew that basket is "dax", they pick the image of a table with a lamp, thereby implicitly learning a match between a referent and a new pseudoword. In total, 20 words were introduced, all of which referred to objects (i.e., pseudowords were nouns). Following the familiarization phase, we administered the test. During the test, participants were shown a sentence (e.g., "Find an image with a dax") followed by a pair of images located in the left and right halves of the screen, respectively. One of these images depicted the correct referent. On each trial, the participant had to indicate on a 4-point scale how confident they were in the correctness of their response. A sample screen (Arabic version) is shown in Fig. 2.

In the auditory modality, the procedure was much the same except that participants *heard* the pseudowords and instruction sentences instead of seeing the orthographic representation of them (no reading involves). The sequence of visual and audio experimental parts and the referents used in the visual and auditory parts were counterbalanced across participants and genders within each language group.

To assess metacognition, we use retrospective confidence ratings. Specifically, we evaluated metacognition as the ability of individuals to track accuracy with confidence ratings (assign lower confidence on the trails, in which the probability of making an error is higher). To measure metacognition, we used the signal detection theory framework. Within

Fig. 2 Sample screen for the visual experiment. Left – learning, right – testing

this framework, correct responses (hits and correct rejections) coupled with high confidence levels were treated as metacognitive hits, while correct responses paired with low confidence were treated as metacognitive misses. Conversely, incorrect responses (misses and false alarms) coupled with high versus low confidence were treated as metacognitive false alarms and metacognitive correct rejections, respectively. To overcome the binary logic (recalling our 4-point confidence scale), we can quantify the metacognitive receiver operating characteristic (ROC) curve (Galvin et al., 2003). Meta-ROC can be estimated in d' units (Galvin et al., 2003; Maniscalco & Lau, 2012). This approach yields a pseudo d' measure that optimally fits confidence ratings; in other words, a would-be d' expected if confidence ratings of an individual optimally discriminated between correct and incorrect responses. This estimated pseudo d', referred to as *meta-d'*, is an individual-specific *metacognitive sensitivity*, which estimates the reliability of his/her confidence ratings, or the individual ability to identify cases when the error is more likely. *Meta-d'*, in other words, reflects how well confidence ratings discriminate between correct and incorrect responses (Fleming & Lau, 2014). Importantly, *meta-d'* (*metacognitive sensitivity*) reflects confidence fluctuations between conditions and is not influenced by differing individual tendencies towards over/under-confidence – that is, by the tendency to assign overall higher or lower ratings to one's own decisions (known as *metacognitive bias*) or by individual differences in cognitive performance (which is self-evaluated by individuals via assigning confidence ratings).

Fleming (2017) developed a method for the hierarchical Bayesian estimation of metacognitive sensitivity and efficiency, which we have adopted for our analysis. M*etacognitive efficiency* is the ratio of *meta-d'* to *d'*, or metacognitive sensitivity given a particular level of cognitive performance, which might be useful in cases when the number of trials across two conditions to be compared varies, hence providing different amount of signal (e.g., Polyanskaya et al., 2022 – see Fig. 1), or when the researcher has theoretical grounds to believe that the task in one condition in inherently more cognitively challenging than in the other condition. For modelling and estimating *meta-d'*, we used the code developed by Fleming (2017; available at https://github.com/metacoglab/Hmeta-d, last verified on April 17, 2023). The data was restructured to fit a stimulus (2) by response (2) by 4 (confidence) matrix, for which we adopted a previously used algorithm (Ordin & Polyanskaya, 2021). Participants were asked to select a referent image, positioned either on the left- or right-hand side of the screen, for the pseudowords. If the correct image was on the left, correct responses were defined as hits and incorrect responses were defined as misses. Conversely,

if the correct image was on the right, incorrect responses were defined as false alarms and correct responses were defined as correct rejections. We then calculated the number of responses for each category (hits, false alarms, correct rejections, misses) at each confidence level, and used that to build the vectors that were used as inputs for the modelling code. Meta-d' and d' measures were estimated separately for each language group.

Results

All tests were first run with *gender* as a factor, but neither the simple effect of gender nor the interaction of gender with other factors were significant. Therefore, in this section we report the tests without *gender* as a factor.

Considering that metacognition is the ability to monitor cognitive performance, and that metacognitive sensitivity (aka, metacognitive accuracy) is the strength of the confidence-accuracy correlation, we first report how well the participants learnt the words. We explored the effect of *group* (Chinese vs. Portuguese vs. Arabic) as a between-subject factor and perceptual *modality* (visual vs. auditory) as a within-subject factor on *accuracy* (i.e., cognitive performance, measured by the number of correct responses). The effect of group, $F(2,223) = 6.011$, $p = 0.003$, $\eta^2_p = 0.051$ and the *group*modality* interaction, $F(2,223) = 7.49$, $p < 0.001$, $\eta^2_p = 0.063$ were significant. To split the interaction, we investigated the main effect of *modality*. This was significant for Arabic speakers, $F(1,83) = 7.091$, $p = 0.009$, who scored better in the auditory modality, and for Portuguese speakers, who scored better in the visual modality, $F(1,71) = 14.694$, $p = 0.012$. Separate ANOVAs were run to probe significance of differences in each modality. The tests revealed that differences in *accuracy* between the groups were not significant in the auditory modality $F(2,223) = 1.329$, $p = 0.267$. In the visual modality, however, the differences were significant, $F(2,223) = 11.786$, $p < 0.001$, $\eta^2_p = 0.096$. Performance of Arabic speakers was lower than that of Chinese speakers, $t = 3.78$, $p < 0.001$, $d = 0.612$ (Cohen's d), and lower than that of Portuguese speakers, $t = 4.44$, $p < 0.001$, $d = 0.713$. There was no significant difference in performance between Chinese and Portuguese speakers in the visual modality, $p = 0.547$ (p-values were corrected for multiple comparisons using the Bonferroni-Holm method). The pattern of results is presented in Fig. 3a.

Second, we analysed metacognitive monitoring (measured as meta-d'). We introduced *modality* as a within-subject factor and *group* as a between-subject factor, with *meta-d'* as a dependent variable. The effect of group, $F(2,223) = 6.888$, $p = 0.001$, $\eta^2_p = 0.058$ and the *group*modality* interaction, $F(2,223) = 15.342$, $p < 0.001$, $\eta^2_p = 0.121$ were significant. To split the interaction, we explored the main effect of *modality*, which was significant for Arabic speakers, who showed better metacognition in the auditory modality, $F(1,83) = 18.613$, $p < 0.001$, and for Chinese speakers, who demonstrated better metacognition in the visual modality, $F(1,69) = 7.476$, $p = 0.008$. Portuguese speakers did not reveal significant differences in metacognition between the modalities, $p = 0.506$. Further, we ran two separate ANOVAs – one for the visual and the other for the auditory modality – and found that the differences between linguistic groups were not significant in the auditory modality, $F(2,223) = 2.144$, $p = 0.12$, with no significant differences between groups pairwise. In the visual modality, the effect of *group* was significant, $F(2,132) = 15.673$, $p < 0.001$, $\eta^2_p = 0.123$ (as the assumption of homogeneity was violated, *df* was corrected using the Welch method). Pairwise comparisons showed that in the visual modality, Chinese exhibited higher metacognitive sensitivity than Arabic, $t = 4.52$, $p < 0.001$ and
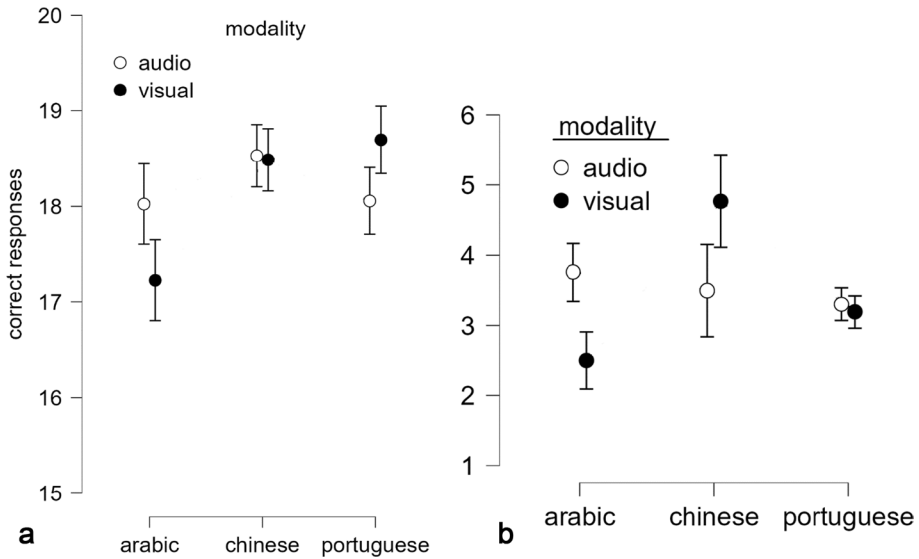
Fig. 3 **a** Number of correct responses per modality and language group. Maximum number of trials is 20 (chance level is 10 correct responses). Error bars stand for 95%CI. **b** Sensitivity of metacognitive monitoring measured as meta-d', per modality and language group. Error bars stand for 95%CI

Portuguese speakers, t = 3.118, p = 0.007 speakers. Portuguese speakers exhibited higher meta-sensitivity than Arabic speakers, t = 2.84, p = 0.014 (as the homogeneity of variance is not assumed, pairwise comparisons were done using the Games-Howell method because homogeneity of variance could not be assumed in all tests, and p values were corrected for multiple comparisons using the Tukey method). The pattern of results is presented in Fig. 3b.

Discussion

The data showed that inferential word learning is more effective in the Arabic population in the auditory modality, and in the Portuguese population in the visual modality, with no significant differences between the modalities in the Chinese population (also confirmed by the 2-tailed Bayesian t-tests, $BF_{10} = 0.151$, indicating that the no-difference hypothesis is 6.63 times more likely than the alternative hypothesis). Performance differences between groups were detected only in the visual modality, with Chinese and Portuguese outperforming Arabic participants.

There was no effect of language on metacognitive sensitivity in the auditory modality. However, in the visual modality, the highest meta-sensitivity was observed in the Chinese group and the lowest in the Arabic group. Arabic and Chinese speakers exhibited higher sensitivity in the auditory and visual modalities, respectively, with no difference in meta-sensitivity between modalities in Portuguese group. Evidence of absence of across-modality differences in metacognition in the Portuguese sample was confirmed by the 2-tailed Bayesian t-tests, $BF_{10} = 0.152$, indicating that the no-difference hypothesis was 6.58 times more likely than the alternative hypothesis.

There is a significant and strong disordinal interaction between *linguistic group* and *perceptual modality*. Namely, (1) metacognition is higher in the auditory modality than in the visual modality among Arabic speakers, whereas the pattern is reversed for Chinese speakers. (2) Arabic speakers show higher rates of learning success in the auditory modality than in the visual modality. This disordinal interaction becomes clearer when we consider that, from infancy onwards, vocabulary expansion during early language acquisition relies predominantly on auditory inferential word learning rather than visual learning. This pattern holds true regardless of the ambient language. Typically developing individuals across all linguistic populations are equally trained in this task and its monitoring, Monitoring is probably trained via their continuous verification of inferences about the semantic meaning of lexical constituents, and their continual updating and refining of these inferences. Consequently, there are no significant differences in the auditory modality between language groups. Because of its lack of variation across languages, we propose using individual performance in the *auditory* modality as a baseline performance level. By contrast, the processing of written information emerges from inculturation at a later age, and it may vary across linguistic populations depending on writing systems and socio-cultural practices, especially those related to educational practices, which in turn may raise or lower word learning performance and its metacognitive monitoring relative to the baseline performance level. Our discussion will focus on the more subtle ways that writing systems can influence metacognition.

With its alphabetic writing and relatively transparent orthography, where symbols systematically and directly correspond to phonemes, Portuguese allows for equal metacognitive sensitivity across modalities. In Arabic, on the other hand, this direct transfer of metacognitive skills to the visual modality is precluded by the opaqueness of the *abjad* alphabet, with each discrete symbol representing a consonant. Vowels are cued by the diacritics that accompany the consonantal symbolic representations; however, not all vowels are cued, nor are vowels always cued, and the same diacritic can be used for other purposes (e.g., as word separators when two consecutive words have the same final and initial consonantal symbols). Additionally, sometimes vowels must be inferred based on grammatical meaning (e.g., verb tense) and at the same time the vowels express the grammatical meanings. We suggest that the presence of symbol-to-sound mapping both in Arabic and Portuguese encourages the transfer across modalities, but that the opacity of the Arabic writing system hinders efficient meta-monitoring of the cognitive processes that match sound with symbolic streams. In the case of Arabic speakers, we see lower metacognitive sensitivity in the visual modality relative to the auditory baseline. By contrast, among the Portuguese population metacognitive sensitivity remains at the baseline level. This might be one of the reasons for the observed disordinal interaction between *linguistic group* and *perceptual modality*.

Another factor contributing to the disordinal interaction is heightened metacognition in the visual modality in Chinese population, surpassing the baseline level. In the simplified Chinese logographic script, the sound-to-symbol mapping is not present. The phonetic radicals provide unsystematic cues to several valid pronunciations for the whole word. Among the Chinese population, we suggest, metacognition in the visual modality tracks the cognitive processes underlying memory, rather than phonological awareness. Tracking memory processes in the visual modality is probably less affected by metacognitive sensitivity in the auditory modality compared to tracking phonological awareness. In other words, the nature of the logographic Chinese script does not provide conditions for the transfer of metacognitive skills from the auditory to the visual modality. The semantic radical of the Chinese characters cues the referent, which must be deduced from previously learned characters, the

context, and the symbol's iconicity. To infer the reference of any given character, therefore, Chinese speakers need to monitor multiple strategies. They must also monitor multiple sets of cognitive mechanisms to simultaneously infer both the pronunciation and the semantics of each holistic character. By contrast, the symbols in classical alphabets or abjads do not have referents that need to be inferred and only refer to phonemic representations (speech sounds), not even to pronunciation of the whole lexical unit.

Overall, our hypothesis was confirmed: languages with different writing systems are associated with differences in metacognition in the word-learning task between the respective populations of speakers in the visual modality, but not in the auditory modality. This finding aligns with our general prediction. Yet our study also yielded further nuances, which underscore the complexity of the association. First of all, we observed a higher metacognitive sensitivity in the visual modality among the Portuguese speakers than among the Arabic ones. This confirmed our specific prediction that orthographic opacity (in this case, the opacity of written Arabic) would impede metacognitive monitoring. We also observed the highest metacognitive sensitivity among Chinese speakers, which aligned with our second specific prediction. Based on these results, we surmise that a wider range of cognitive processes pertaining to memory and phonological awareness – processes that we assume to be implicated during processing of written information in a logographic Chinese script – lead to greater metacognitive enhancement. Metacognition among the Chinese population, we propose, is further trained by the need to simultaneously monitor cognitive strategies for both pronunciation and semantics while processing a single character. This result aligns with the theory proposed by Polyanskaya et al. (2022), which posits that bilinguals with typologically different languages should exhibit enhanced metacognition due to the need to monitor two sets of processing strategies, one for each language in their inventory.

The presence and transparency of sound-to-symbol mapping influences the success of word acquisition and the self-evaluation (metacognition) of word learning, when words are visually presented. Although some metacognitive processes take place in the absence of conscious awareness (Jachs et al., 2015; Kentridge & Heywood, 2000; Kunimoto et al., 2001; Pournaghdali et al., 2023), metacognition and conscious awareness of learning content and processes are interrelated (Kentridge & Heywood, 2000; Ordin & Polyanskaya, 2021; Schraw, 1998), and metacognitive sensitivity can be used as a valid proxy to study how learning success and conscious awareness of what is being learnt are interacting (Ko & Lau, 2012; Maniscalco & Lau, 2012; Persaud et al., 2007; Shimamura, 2008). Our study focused only on metacognitive processes that are related to conscious awareness because we assessed metacognition by eliciting retrospective confidence ratings of participants about correctness of each of their responses, which required participants to consciously self-evaluate their cognitive performance. Evidence suggests that awareness – cued by metacognitive sensitivity in the task that requires people to consciously monitor the learning outcomes – may also influence learning success (Jacob, 1998; Soto et al., 2023; Stanton et al., 2021; Taouki et al., 2022). This is particularly relevant in explicit learning scenarios such as formal language learning, where consciousness plays a role in vocabulary building, and where learning success is modulated by the individual's conscious awareness of what is being learnt and the effectiveness of the learning process (Ordin & Polyanskaya, 2021). Such consciousness comes into play when a word is associated with an object: when words are represented visually, differences in one's conscious awareness of success in learning the words make differential demands on word learning across languages and modalities. The differences we observed among users of different writing systems and between perceptual modalities may potentially modulate the effectiveness of vocabulary acquisition activities during foreign language learning.

Given the observed impact of writing systems on metacognitive sensitivity in language learning, several avenues for further research emerge. One potentially fruitful approach would be to compare the relative effectiveness of different teaching activities and methods in foreign language vocabulary building, taking into account the writing scripts in the native and target languages of learners. Providing auditory support may well enhance students' learning of written words in Arabic, while visual support aids may have the same effect on spoken word learning in alphabetic languages, ultimately improve students' pronunciation and expediting vocabulary expansion. The pedagogical implications of our results pave the entrance into our future research agenda, which will aim to identify how vocabulary-building activities can be adapted to the writing systems of specific target languages.

Conclusion

We explored how experience with different writing systems modulates metacognition during language tasks that involve versus those do not involve the processing of written information. To this end, we administered an inferential word learning task in the auditory and visual modalities to native speakers of the following languages: Portuguese (alphabetic script, transparent orthography), Standard Arabic (quasi-syllabic abjad – where the symbol representing a consonant may sometimes have diacritics representing the following vowel, opaque orthography), and Mandarin Chinese (logographic script, no phoneme-to-symbol mapping). In the auditory modality, no differences between speakers of these languages were observed – neither in terms of word learning performance, nor in the ability to evaluate how well the words were learnt (metacognition). Conversely, in the visual modality, both learning and self-evaluation differed between language groups, suggesting that a person's writing system affects word learning in the visual but not in the auditory modality. Specifically, we observed lower metacognition in the visual modality in the Arabic than in the Portuguese speaking population, which we explained by the relative opacity of Arabic orthography. Despite the absence of phoneme-to-symbol mapping in the Chinese script, metacognition in the visual modality in the Chinese population was the highest because of the need to monitor a wider range of cognitive processes and strategies during processing written information. The observed differences suggest that the activities for vocabulary learning in formal classroom should be adapted to the writing system in the target language.

Limitations

(1) Metacognition can be affected by a variety of factors including genetics (Greven et al., 2009), educational practices (de Jager et al., 2005; Yeager et al., 2019), and short-term mental states (Nicholson et al., 2019; Rouault et al., 2018). These factors might potentially contribute to differences in metacognition between populations in diverse societies across the globe. Moreover, both theoretical (e.g., Dunstone & Caldwell, 2018; Heyes et al., 2020) and empirical (e.g., van der Plas et al., 2022) studies point to evidence for the cultural impact on metacognition, suggesting that the observed differences in metacognition between Chinese, Portuguese and Arabic populations may be cultural rather than linguistic. For example, successful inter-personal communication

and collaborative activities require a nuanced understanding of others' intentions and mental states (Bahrami et al., 2010; Fusaroli et al., 2012; Reddish et al., 2013). This awareness of others' intentions, in turn may enhance an individual's meta-monitoring skills (Heyes et al., 2020). As societal perceptions of the importance of interpersonal understanding and social collaboration vary significantly (Cleeremans et al., 2020; Markus & Kitayama, 2010), the precision with which individuals are able to understand others' mental states and intentions also varies across the globe (Heyes & Frith, 2014). This might underly some culturally-determined differences in metacognition across populations; in other words, we cannot exclude the cultural origin of the observed effect. Nevertheless, the cultural variations could influence differences in metacognition *between* groups across both perceptual modalities, but not *within* groups, between visual and auditory modalities. We observed the differences in metacognition in the visual modality, but not in the auditory modality, as we had predicted. It is difficult to explain why cultural influence would be observed in one modality but not in the other, while differences in writing scripts may account for whole result pattern. Despite the plausibility of the alternative interpretation, the nuances of the result pattern support the idea that differences between writing systems play a more crucial role in modulating metacognition in the word learning task. Therefore, we advocate the linguistic rather than cultural origin of the observed differences in metacognition.

(2) Data in Saudi Arabia was collected separately on men's and women's campuses, because of the sex-based segregation policy in the higher education sector. Therefore, only male researchers worked with male Arabic participants, and female researchers – with female Arabic participants. In Portugal and in China, male and female research assistants worked with both genders, and the campuses are mixed. Although we doubt that it might have influenced the results, we do not have empirical evidence to exclude this possibility. Again, we do not think that segregation policy might have influenced metacognition in the visual modality relative to the auditory modality, or metacognition only in visual but not in auditory modality.

(3) As we wrote earlier, retrospective confidence judgments may separate between correct and incorrect responses without conscious awareness (Jachs et al., 2015; Pournaghdali et al., 2023). We admit that we cannot separate between the contribution of confidence judgments guided with and without conscious awareness. Importantly, confidence ratings are guided without conscious awareness when the accuracy is at or close to the chance level (because there is little signal that is consciously processed). We believe that in our study the contribution of consciousness into confidence ratings is greater than contribution of whatever was happening outside conscious awareness, because cognitive performance was relatively high for all individual participants, and participants could consciously evaluate their performance.

Data availability Following the data management plan of the ERC project 101040850 – TypaMetaLing – the data will be made publicly available upon publication on the FigShare repository associated with the project, data on participants will be anonymized. If reviewers need to access the data during the evaluation process, the data can be made available – privately – by sharing this link: https://figshare.com/s/38e6309dfb 1a5c504183.

Declarations

Ethics approval The experimental protocol was approved by the ethical review board of the Prince Sultan University (Saudi Arabia, El Riyadh) and of the School of Psychology and Educational Sciences, Coimbra University (Coimbra, Portugal). ERB of SPES (Coimbra) made sure that the study conformed to the recognized standards of the Declaration of Helsinki. ERB at the PSU made sure that the stimuli are culturally suitable in the Saudi society.

Competing interests We have no conflict of interest to disclosure.

References

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science, 329*(5995), 1081–1085.

Bolger, D., et al. (2005). Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping, 25*(1), 92–104.

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General, 148*(1), 51–64.

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., Vuillaume, L., & de Heering, A. (2020). Learning to be conscious. *Trends in Cognitive Sciences, 24*(2), 112–123.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24–39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science, 17*, 905–912.

De Jager, B., Jansen, M., & Reezigt, G. (2005). The development of metacognition in primary school learning environments. *School Effectiveness and School Improvement, 16*(2), 179–196.

Dong, J., Li, A., Chen, C., Qu, J., Jiang, N., Sun, Y., Hu, L., & Mei, L. (2021). Language distance in orthographic transparency affects cross-language pattern similarity between native and non-native languages. *Human Brain Mapping, 42*(4), 893–907.

Dunstone, J., & Caldwell, C. A. (2018). Cumulative culture and explicit metacognition: A review of theories, evidence and key predictions. *Palgrave Communications, 4*, 145. https://doi.org/10.1057/s41599-018-0200-y

Ferrand, L. (2007). *Psychologie cognitive de la lecture. Reconnaissance des mots écrits chez l'adulte. Ouvertures psychologiques*. Bruxelles: De Boeck.

Fleming, S., & Lau, H. (2014). How to measure metacognition. *Frontiers in Psychology, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness, 2017*(1):nix007. doi: 10.1093/nc/nix007

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology, 75*, 80–96.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science, 23*(8), 931–939.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin and Review, 10*(4), 843–876.

Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., & Plomin, R. (2009). More than just IQ. *Psychological Science, 20*(6), 753–762.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091.

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences, 24*(5), 349–362.

Jachs, B., Blanco, M., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 269–276.

Jacob, P. (1998). Memory, learning and metacognition. *Comptes Rendus De L'académie Des Sciences - Series III - Sciences De La Vie, 321*(2–3), 253–259.

Kentridge, R. W., & Heywood, C. A. (2000). Metacognition and awareness. *Consciousness and Cognition, 9*, 308–312.

Ko, Y., & Lau, H. (2012). A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philosophical Transactions of the Royal Society b: Biological Sciences, 367*, 1401–1411.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10*(3), 294–340.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430.

Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science, 5*(4), 420–430.

New, B., Bourgin, J., Barra, J., & Pallier, C. (2024). UniPseudo: A universal pseudoword generator. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218231164373

Nicholson, T., Williams, D. M., Grainger, C., Lind, S. E., & Carruthers, P. (2019). Relationships between implicit and explicit uncertainty monitoring and mindreading: Evidence from autism spectrum disorder. *Consciousness and Cognition, 70*, 11–24.

Ordin, M., & Polyanskaya, L. (2021). The role of metacognition in recognition of the content of statistical learning. *Psychonomic Bulletin and Review, 28*(1), 333–340.

Ordin, M., Polyanskaya, L., & Soto, D. (2020). Metacognitive processing in language learning tasks is affected by bilingualism. *Journal of Experimental Psychology: Learning Memory and Cognition, 46*(3), 529–538.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*, 257–261.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*(2), 201–236.

Polyanskaya, L. (2022). Cognitive mechanisms of statistical learning and segmentation of continuous sensory input. *Memory and Cognition, 50*, 979–996.

Polyanskaya, L., Manrique, H. M., Marín, A., García-Palacios, A., & Ordin, M. (2022). Typological differences influence the bilingual advantage in metacognitive processing. *Journal of Experimental Psychology: General, 151*(11), 2706–2719.

Pournaghdali, A., Schwartz, B. L., Hays, J., & Soto, F. (2023). Sensitivity vs. awareness curve: A novel model-based analysis to uncover the processes underlying nonconscious perception. *Psychonomic Bulletin and Review, 30*, 553–563.

Quine, W. (1960). *Word and object*. MIT Press.

Rademaker, R. L., & Pearson, J. (2012). Training visual imagery: Improvements of metacognition, but not imagery strength. *Frontiers in Psychology*, *3*(JUL), 224. https://doi.org/10.3389/fpsyg.2012.00224

Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's dance together: Synchrony, shared intentionality and cooperation. *PLoS ONE, 8*, e71182. https://doi.org/10.1371/journal.pone.0071182

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry, 84*(6), 443–451.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125.

Seymour, P., et al. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174.

Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 373–390). Psychology Press.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating interindividual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition, 20*(4), 1787–1792.

Soto, C., Gutierrez de Blume, A. P., Rebolledo, V., et al. (2023). Metacognitive monitoring skills of reading comprehension and writing between proficient and poor readers. *Metacognition and Learning, 18*, 113–134.

Stanton, J.D., Sebesta, A.J,. Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE-Life Scinces Education*, *20*(2):fe3, 1–7. https://doi.org/10.1187/cbe.20-12-0289

Taouki, I., Lallier, M., & Soto, D. (2022). The role of metacognition in monitoring performance and regulating learning in early readers. *Metacognition and Learning, 17*(3), 921–948.

Thiessen, E., Kronstein, A., & Hufnagle, D. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*(4), 792–814.

Thorstad, G. (1991). The effect of orthography on the acquisition of literacy skills. *British Journal of Psychology, 82*(4), 527–537.

Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, Y.-W.F., Wang, S., & Chen, H.-C. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods, 50*, 1763–1777.

van der Plas, E., Zhang, S., Dong, K., Bang, D., Li, J., Wright, N. D., & Fleming, S. M. (2022). Identifying cultural differences in metacognition. *Journal of Experimental Psychology: General, 151*(12), 3268–3280.

Yeager, D. S., Hanselman, P., Walton, G. M., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*, 364–369.