SHORT TITLE: Bilingualism, typological difference and metacognition

TYPOLOGICAL DIFFERENCES INFLUENCE THE BILINGUAL ADVANTAGE IN METACOGNITIVE PROCESSING

Leona Polyanskaya (*), Department of Psychology and Sociology, University of Zaragoza, Spain; Slovak Academy of Sciences, Bratislava, Slovakia

Héctor M. Manrique, Department of Psychology and Sociology, University of Zaragoza, Spain

Antonio Marín, Department of Basic Psychology, Clinical and Psychobiology, Jaume I University, Castellon, Spain

Azucena García-Palacios, Department of Basic Psychology, Clinical and Psychobiology, Jaume I University, Castellon, Spain; CIBER Fisiopatología Obesidad y Nutrición (CIBERObn), Instituto Carlos III, Madrid, Spain.

Mikhail Ordin, BCBL, San Sebastian, Spain; University of Konstanz, Linguistics Department, Konstanz, Germany

(*) corresponding author

Address for correspondence: leona.polyanskaya@gmail.com

Abstract:

Previous studies showed a bilingual advantage in metacognitive processing (tracking one's own cognitive performance) in linguistic tasks. However, bilinguals do not constitute a homogeneous population, and it was unclear which aspects of bilingualism affect metacognition. In this project, we tested the hypothesis that simultaneous acquisition and use of typologically different languages leads to development of diverse processing strategies and enhances metacognition. The hypothesis was tested in the visual and auditory modalities in language and non-language domains, in an artificial language learning task. In the auditory modality, the hypothesis was confirmed for linguistic stimuli, with no between-domain transfer of metacognitive abilities was observed at the individual level. In the visual modality, no differences in metacognitive efficiency were observed. Moreover, we found that bilingualism per se and the use of typologically different languages modulated separate metacognitive processes engaged in monitoring cognitive performance in statistical learning task.

Introduction:

Living beings continuously make decisions, evaluate the efficacy of these decisions, and adapt their behavior accordingly. Animals – including humans – evaluate their own cognitive processes, the available evidence, as well as the degree of uncertainty associated with decisions taken based on this evidence (Kepecs et al., 2008; Smith 2009). The ability to monitor and regulate one's own cognitive processes and behavior is referred to as metacognition. Metacognition has two important components: knowledge about cognition and regulation of cognition, which in turn affords regulation of behavior (Flavell, 1979; Schraw, 1998). These components are served by different cognitive processes, referred to as *metacognitive monitoring* and *metacognitive control* (Dunlosky et al., 2007; Nelson & Narens, 1990). Monitoring processes track decisions, cognitive states, and behavior in uncertain situations and underpin retrospective estimates of confidence associated with cognitive states and past decisions (Flavel, 1979; Kepecs et al., 2008). Control processes are deployed to guide future behavior given one's current cognitive states, available evidence on the current environment, and past outcomes. Individuals with good metacognitive monitoring abilities assign higher confidence to decisions which are less likely to be mistaken; this means confidence estimates help discriminate between correct and incorrect decisions. This ability to discriminate on the basis of confidence is often assumed to rely on conscious awareness of stimuli (Persaud et al., 2007; Smith et al., 2003), error monitoring and detection mechanisms (Ordin et al., 2020), and evaluating one's own cognitive states (Smith, 2009).

Notably, metacognition is a separate ability from cognition. It is possible for an individual to perform a cognitive task well, but fail to realize that his performance is good; this might be reflected in under-confidence in his decisions. It is also possible for an individual whose performance is below average to consider his level of performance is high, that is be overconfident about his decisions. In both cases, individuals fail to monitor and adequately evaluate their cognitive performance, that is, reveal low metacognitive ability, irrespective of their cognitive performance. Metacognitive ability allows an individual to be more confident in his decisions when cognitive performance is optimal. Studies on bilingualism and cognition have largely focused on the modulatory effect of linguistic experience on executive functions, attention, memory (Abutalebi & Green, 2007;

Bialystok et al., 2004; Bialystok & Viswanathan, 2009; Blumenfeld & Marian, 2014; Carlson & Meltzoff, 2008; Costa et al., 2009; etc.). To date, only two studies have addressed the effects of linguistic experience on monitoring cognitive performance (Folke et al., 2016; Ordin et al., 2020), although it is a distinct process that may be efficient or inefficient irrespective of how good cognitive performance is. However, it is known that executive functions can modulate cognitive control and decision-making (Del Missier et al., 2010; Fernandez-Duque et al., 2000), that decision-making is also related to retrospective confidence (confidence in already-made decisions) and the level of uncertainty about available evidence (Kepecs et al., 2008), meaning that executive functions are potentially related to metacognitive monitoring via decision-making. As bilingual experience can modulate executive functions and cognitive control, it could also modulate metacognitive monitoring.

Some recent evidence suggests that metacognitive skills can be modulated by individual experience in a particular task or operational domain (Carpenter et al., 2019; Rademaker & Pearson, 2012). Ordin et al. (2020) showed that bilinguals' wider experience with a rich range of linguistic cues and language structures leads to enhanced metacognitive monitoring in statistical learning tasks. Performance on these tasks is known to correlate with general language abilities (Erickson & Thiessen, 2015; Saffran, 2002; Siegelman, 2020) and underpin efficient speech and language processing (Kahta & Schiff, 2019; Kidd, 2012; Misyak & Christiansen, 2012; Saffran, 2018).

Importantly, bilinguals do not constitute a homogeneous population. Bilinguals differ on many dimensions: they may speak several languages from birth or have acquired them sequentially; they may be equally proficient in multiple languages or predominantly use one of their languages; they may have literacy skills in one language but only oral competence in another language; they may mix languages within the same communicative context or use them separately in different social settings (e.g., use one language in a professional context and another language at home). All of these factors are important when investigating how bilingualism might lead to metacognitive enhancement in language tasks. For example, exposure to a bilingual environment and simultaneous acquisition of multiple languages in early life can influence the development of neural and cognitive processes and neural architecture in the course of individual development (Birdson, 2009; 2018), while maturational constraints after the so-called "critical period" may prevent neural rewiring in those brain substrates and networks that underlie metacognition (Johnson & Newport, 1989; Long, 1990). An alternative hypothesis is that it is not early exposure to multiple languages but rather proficiency in multiple languages that leads to metacognitive enhancement. In the former case, early bilinguals should outperform late bilinguals in metacognitive processing even if they lost one of their languages later in life, because neural rewiring already took place under the pressure of a bilingual environment at an early age. In the latter case, even L2 learners who have acquired another language in adulthood will outperform early bilinguals who have lost one of their languages, because they need to constantly monitor language use and apply language control processes (Abutalebi & Green, 2007; Woll, 2018). Determining which of these two hypotheses is true requires targeted investigation and falls beyond the scope of this study. But these questions illustrate why age of acquisition and balanced proficiency, for example, are bilingualism-related factors that may affect how individual linguistic experience modulates metacognition.

Additionally, the languages in a bilingual's inventory might be typologically similar, allowing language structures of both languages to be processed by a similar set of cognitive processing strategies, or typologically different, requiring specialized cognitive strategies adapted for processing distinct linguistic structures. Empirical evidence for a bilingual advantage in metacognitive monitoring of language tasks stems from comparing Basque-Spanish bilinguals and Spanish monolinguals (Ordin, Polyanskaya, & Soto, 2020). However, the superior performance of Basque-Spanish bilinguals might not result from bilingualism per se but rather the use of two typologically different languages: Basque, a non-Indo-European language, and Spanish differ in terms of canonical word order (SVO vs. SOV), alignment of core arguments (ergative vs. accusative), acoustic manifestation, and phonological placement of stress. These typological differences might engage different processing strategies, leading to enhanced metacognitive monitoring by Basque-Spanish bilinguals in language tasks. Here, we ran a hypothesis-driven study that aimed to test whether individual experience with typologically different languages facilitates metacognitive enhancement in statistical learning tasks, controlling for other bilingualism-related confounding factors (age of acquisition, IQ, relative proficiency in both languages, constant immersion into bilinguals environment on a daily basis).

Statistical learning operates across operational domains, perceptual modalities, and types of material (Baldwin et al., 2008; Conway, 2020; Conway & Christiansen, 2005; 2006; 2009; Frost et al., 2015; Gebhart et al., 2009; Hard et al., 2011). This leads us to ask whether statistical learning is domain- and/or modality-specific, and whether it consists of a single mechanism, or a set of mechanisms (Siegelman et al., 2017; Thiessen et al., 2013), with different subsets being engaged depending on the domain, modality, and the nature of the material being processed. Testing for within-subject differences in metacognitive efficiency could help us pinpoint whether metacognitive processes track the same or different sets of cognitive mechanisms when statistical learning tasks are performed on linguistic versus non-linguistic material in visual versus auditory modalities. We also explore limits to the transferability of metacognitive skills between language and non-language domains.

To address our research interests, we compared metacognitive efficiency in an artificial language learning task performed by Spanish monolinguals and two groups of bilinguals – Spanish-Catalan (typologically similar languages) and Spanish-Basque (typologically different languages) – controlling for factors related to relative proficiency in native and foreign languages, language use, as well as verbal and logical IQ. Typological similarity is measured as the number of structural properties shared by two languages. Catalan and Spanish are strongly suffixing languages (they do not use prefixes to express grammatical meanings), exhibit accusative alignment of core elements, have SVO canonical word order, and use prepositions. Basque is a morphologically balanced language that uses both prefixes (although, this is a non-productive or low-productive feature of modern Basque) and suffixes to express grammatical meaning, ergative alignment, canonical SOV word order and postpositions, including articles attached to the end of noun phrases. Conjunctions that express relations in sentences with subordinate clauses are typically placed after the subordinate clause in Basque, while in Spanish and Catalan, they are placed before the subordinate clause. In sum, Catalan and Spanish are typologically more similar than Basque and Spanish.

The main statistical learning task was administered in two perceptual modalities (visual and auditory), both on linguistic (sequences of syllables) and non-linguistic (sequences of environmental sounds or fractal images) material, in a 2*2*3 experimental design (*modality* and *domain* as within-subject factors and *group* as a between-subject factor. We predicted that Basque-Spanish bilinguals would exhibit higher metacognitive efficiency than both Spanish monolinguals and Catalan-Spanish bilinguals. Metacognitive efficiency was measured using a signal detection analytic approach and Bayesian hierarchical modelling (Fleming, 2017).

Methods:

The project was approved by the ethical board at the BCBL (approval number 06092019ML2).

Experimental Procedure

We adapted a classical artificial language learning experiment (Saffran, Aslin, & Newport, 1996) to address the objectives of our study. In the first phase of the experiment, participants were exposed to continuous sensory input composed of recurrent constituents. No sensory cues indicated the edges of these constituents yet, over the course of exposure, participants were able to detect structural regularities. This is a well-known effect in statistical learning, where regularities help participants detect boundaries between constituents and extract them as discrete units within a continuous sensory flow. After the exposure phase, participants completed a recognition test. During the recognition test, they were presented with a short stimulus sequence and had to indicate whether this sequence had been a discrete constituent in the exposure stream (in a different version of the test, participants might be presented with several stimulus sequences, and asked to choose which of them was a discrete constituent from the exposure stream, yet in this experiment we preferred a yes/no alternative of a recognition test for more straightforward application of the signal detection analysis – see below). On each trial, we additionally collected participants' confidence ratings; they indicated how sure they were of their response on a 4-point scale.

The experiment was carried out on linguistic and non-linguistic material in the visual and auditory modalities. Before attending the linguistic material, participants were informed that they were going to listen to alien speech (in the auditory modality) or see a text in an alien language (in the visual modality); they should try to detect the words in that language. Before attending the non-linguistic material, participants were asked to detect recurrent sequences of sounds (auditory modality) or images (visual modality). Although statistical learning is often treated as a type of incidental learning, in this experiment two key considerations motivated the use of explicit instructions. First, we used a within-subject approach because we wanted to directly compare how typological differences between the languages in a bilingual's inventory would affect metacognitive efficiency across domains and modalities. This meant that each participant performed the experiment in both modalities and on both types of material. Thus, if participants had not received explicit instructions, they would nevertheless have been aware that the task was followed by a test when they started their second session; that is, regardless of the initial instructions, participants would be soon find themselves in an explicit learning situation. Second, empirical evidence shows that statistical learning engages both implicit and intentional learning processes (Ordin & Polyanskaya, 2021; Turk-Browne et al., 2005). Explicit instructions turn incidental into intentional learning (Arciuli et al., 2014; Reber et al., 1980; 1991; Schiff et al., 2017), and differences in metacognitive efficiency emerge to a greater extent in explicit learning situations (Schraw, 1998). Since the focus of this study was the modulatory effect of linguistic typology on metacognition, which is related to conscious awareness (Kunimoto et al., 2001; Nelson, 1996; Persaud et al., 2011; Persaud et al., 2007; Shimamura, 2008), explicit instructions make more sense in the context of this study.

Each participant came to the lab twice, with a one- to two-week interval between sessions. In one session, they performed the experiment with linguistic materials (first in one modality, then the other), and in the other session they were exposed to the non-linguistic material. The order of sessions and the order of modalities within sessions was counterbalanced across participants. After the final session, all participants performed IQ tests (described below), a rapid picture naming task, and filled in a language-use questionnaire to elicit information about the age of acquisition of their language(s), percentage of time each language was used in various social contexts, their level of proficiency in other foreign languages, etc.

Participants

We recruited native Spanish monolingual speakers residing in the province of Teruel in Aragon (N=39), Catalan-Spanish bilinguals in the province of Castellón in Valencia (N=43), and Basque-Spanish bilinguals from the province of Gipuzcoa in the Basque Country (N=46). They were all students from the University of the Basque Country, Zaragoza University, or Jaume I University. None of the participants reported any speech/language/hearing disorders. All of the bilingual participants reported they were more frequent users of their national minority languages – Catalan or Basque – in both professional and home environments and preferred to use these languages to read and watch news, speak with family members and friends, and, wherever possible, in formal communication settings. Only those bilinguals who were raised by native Catalan or Basque speaking parents were included in the samples. Hence bilingual samples were represented by simultaneous bilinguals and had equivalent AoA. Students of linguistics, modern languages and translation, and individuals who reported being regular users of foreign languages or who had learnt a foreign language beyond the obligatory hours of formal education were excluded from our samples.

Catalan bilinguals were native speakers of Valencian (a variety of the Catalan language used in Castellón), Basque bilinguals were native Batua speakers (the standard variety of the Basque language). All participants were between 18-30 years of age (M=25 for Basque and Catalan bilinguals and M=24 for Spanish monolinguals); 60% were females (sex distribution was the same across groups). Two Catalan speakers were trilinguals (with Romanian as an additional language acquired simultaneously with Catalan and Spanish, from one of their parents in their family environment). Their scores were not different from the other Catalan bilinguals, and since Romanian, as a Romance language (Andreose & Lorenzo, 2013; Eberhard et al., 2020), is typologically similar to Spanish and Catalan, we therefore decided to keep these trilinguals in the sample of bilingual speakers with typologically similar languages.

To assess bilinguals' relative proficiency, we used a picture naming task based on the approach proposed by Gollan et al. (2012). The test probed lexical access, which was used as a proxy for the relative proficiency of bilinguals in their two languages. We selected 65 pictures representing common entities from different categories (animals, body parts, everyday objects), ensured they were non-cognates in both language pairs (e.g., *mesa* (Spanish); *taula* (Catalan); *mahaia* (Basque); gloss: *table*). Bilinguals named all the objects first in one language, then in the other language (the order of languages was counter-balanced across participants; instructions were given in the corresponding language). In each language, for each picture that was correctly named, participants received one point. All bilinguals achieved the maximum score – 65 points – in the majority language, Spanish. Basque bilinguals (median M=64) scored higher in their minority language than Catalan bilinguals (M=61, Figure 1a) at the group level. In both bilingual groups, the ceiling effect skewed the distribution of scores, so we used Mann-Whitney U-test to statistically compare the scores between bilingual groups, $W=1636.5$, $p<.001$. Although this difference was statistically significant, the difference in median scores was based on only 3 points out of 65. A possible explanation for this small difference is that when Catalan participants did the task in Catalan, they were more likely to substitute a Spanish word than their Basque counterparts, without having noticed to have done so. Since Spanish and Catalan belong to the same language family and share a large number of vocabulary stems, the use of a Spanish word in a Catalan sentence does not sound unnatural, even if there is a corresponding Catalan word. However, using a Spanish word in a Basque sentence may require adapting morphology, for example, moving the article from the Spanish pre- to the Basque post-noun position, or deciding how to assign noun gender in Spanish since Basque does not have gender markers. In other words, the relative ease of substituting a Spanish for a Catalan word led some Catalan-Spanish bilinguals to occasionally assign Spanish names to pictures, which was considered as an incorrect response and resulted in slightly lower scores on the test ($\Delta M=3$). Alternatively, the significant difference might reflect lower proficiency of Catalan than Basque bilinguals in their minority language. Although we do not find it a likely explanation, we cannot exclude this alternative. Potentially, relative proficiency in both languages might also affect bilinguals' metacognitive sensitivity in language tasks, hence we decided to include this score as a covariate in the statistical models (for monolingual participants, we included the picture naming test score in Spanish, which was the ceiling score for all Spanish monolinguals). Additionally, we ran the analyses without covariate, and the results did not change. Here, we report the statistics with the covariate included.

To test cross-sample differences in IQ, we used a Spanish version of a subset of the KBIT2 test (Kaufman & Kaufman, 2004) that yields a normalized score for logical IQ. This KBIT2 test was chosen because of its brevity (compared to a full-range IQ test) and its proven validity and reliability in capturing individual differences in both research and clinical contexts (Scattone, Raggio, & May, 2001; Kievit et al., 2016). We used histograms to verify that the scores were normally distributed in each group and confirmed that the data did not deviate from the normal distribution by means of the Shapiro-Wilk test (p>.02 for each group). Levene's test did not identify significant differences in variance (p=.158). The mean and median (M) scores were close in each group: Basque bilinguals (111.5, SD=8.66, M=110), Catalan bilinguals (110.5, SD=6.83, M=110), and Spanish monolinguals (110.7, SD=7.08, M=112). An Analysis of variance (ANOVA) did not reveal any significant differences between group means, F(2,124)=.209, p=.812 (Figure 1b).

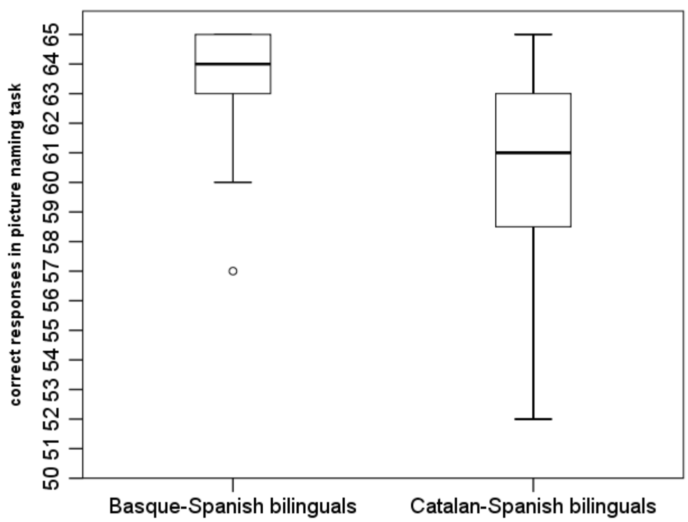

Figure 1a. Number of correct responses in Catalan (for Catalan-Spanish bilinguals) and in Basque (for Basque-Spanish bilinguals) in the picture naming task. The box shows the middle 50% range; the horizontal line inside the box shows the median; the bars represent the first and the fourth quartiles. The range of possible scores is between 0 and 65 (minimum and maximum scores).

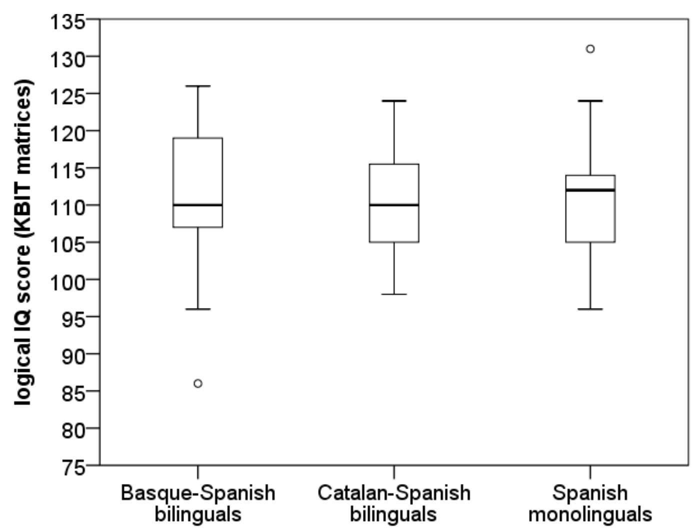

Figure 1b. KBIT logical IQ scores. The box shows the middle 50% range; the horizontal line inside the box shows the median; the bars represent the first and the fourth quartiles. The range of possible scores is between 52 (minimum) and 131 (maximum) points.

Material

We used the linguistic and non-linguistic material from the experiments reported in Ordin et al. (2021; excluding the semi-linguistic material). For the readers' convenience, an abridged description of this material is presented below.

Linguistic input, auditory modality

We selected 32 open (consonant-vowel) syllables that are phonotactically legal and common in all native languages of participants. 24 of these syllables were organized into eight triplets (the “words of an alien language”): /ko-fa-me/, /fo-na-ku/, /mo-si-ke/, /ka-so-ni/, /sa-mu-pe/, /no-su-pi/, /po-fu-mi/, and /fe-nu-pa/. Each of these 24 syllables was used only once (in one of the triplets) so that the transitional probabilities (TPs) between syllables within triplets were all 100%.

Another eight syllables from the inventory – /ma/, /fi/, /pu/, /se/, /ne/, /ki/, /li/, and /lu/ – were used as single-syllable fillers, modeling the use of function words in natural languages. This was a modification of an approach introduced by Gervain et al. (2008), which makes use of frequent syllables to model function words in natural languages. We used eight syllables – /ma/, /fi/, /pu/, /se/, /ne/, /ki/, /li/, and /lu/ – as the functional elements between content words (simulating articles, prepositions, inflexions). TPs between these fillers and the triplet-boundary syllables were approximately 12.5%. The difference in TPs between syllables within triplets and between syllables straddling the triplet boundaries provided a statistical basis for participants to extract the triplets from the input. A Basque-Spanish and a Catalan-Spanish bilinguals checked the triplets and possible triplet-filler concatenations to make sure that real words from participant’s native languages did not emerge in the exposure stream.

Also, we imposed a language-like prosodic hierarchical structure on the syllabic stream, to make the syllabic stream sound more like speech composed of longer sentences. Two triplets with surrounding fillers were organized into phonological phrases (PPs), and two consecutive PPs were grouped into intonational phrases (IPs), or sentences. PPs within the same IP were separated by a short pause, and PPs belonging to different IPs were separated by a longer pause. Overall, each triplet was presented 80 times during exposure, and we avoided repeating the same triplet within one IP prosodic frame. We also imposed an intonational contour, with an overall declination trend, placing boundary rising tones on the PP-initial syllable to indicate the beginning of a phrase and on the PP-final syllable within the IP to indicate that the sentence would continue, and a falling boundary tone on the IP-final syllable to indicate the end of the sentence. This structure is typical of prosodic hierarchy in natural languages (Nespor & Vogel, 2007); this hierarchy is manifested by acoustic cues

(Gussenhoven, 2004; Ladd, 2008), which we implemented in our material (pauses, rising and falling boundary tones, declination of F0 contour over IPs, co-articulations within PPs and breaks of co-articulations at the PP-boundaries). Figure 2a represents the grouping of syllables into PPs and IPs. Figure 2b shows a waveform, and the spectrogram and intonational contour of a PP pair within a single IP. Note that pauses and boundary tones are not aligned with the edges of the triplets and thus cannot be used to detect the triplet boundaries; this prevents participants from using acoustic cues instead of statistical cues to segment recurrent triplets from a continuous syllabic stream.

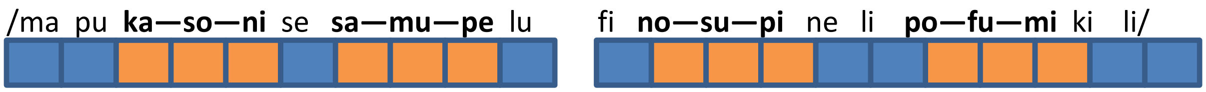

Figure 2a. A schematic representation of a syllabic stream, showing 4 triplets (orange squares, bold font), with fillers (blue squares, normal font), and a 50-ms within-IP between-PP pause.

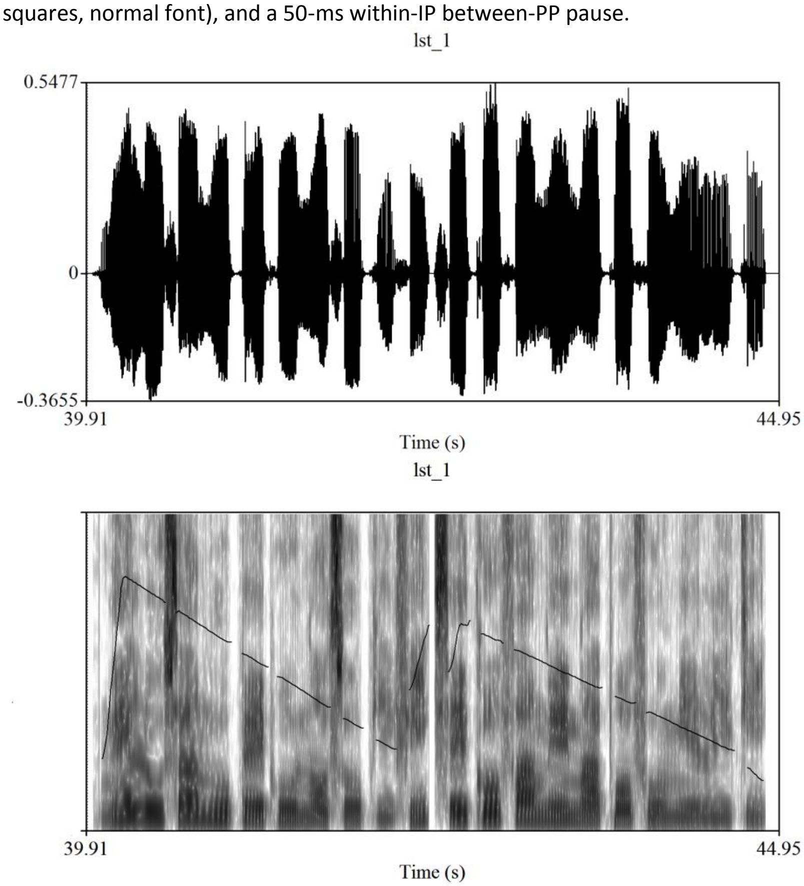

Figure 2b. Waveform and spectrogram representing a single IP, with a 50-ms pause between two PPs, defined by boundary tones. The spectrogram is displayed on a 0 to 7000Hz scale, and the pitch contour is displayed on a 50 to 250Hz scale.

MBROLA (Dutoit & Leich, 1993) was used to synthesize the speech. We used the SP2 voice (Spanish), which produces phonemes in a Spanish-specific manner, making the input more speech-like for our participants, who all had Spanish as a native language. Syllable duration was 240ms (140ms – vowels), with 50ms within-IP pauses and 150ms between-IP pauses. The F0 downtrend went from 210Hz to 100Hz, dropping from 100Hz to 80Hz on IP-final falling boundary tones.

For the recognition test, we created isolated triplets of syllables, monotonized with F0 set to 120Hz. Each syllable was 240ms (with vowels 140ms). Foils were created using the same syllables that were used to compose the triplets, but concatenated such that the consecutive syllables in the foils had never occurred consecutively in the exposure stream (i.e., the TPs between syllables in the foils were 0% in the stream). Non-adjacent TPs (between the first and the third syllables) were also set to 0. In total, 16 foils were created. In eight foils, we preserved the order of syllables (if a particular syllable was used in the triplet-initial position in an exposure stream, it was also placed in the foil-initial position). In the other eight foils, the positional order of syllables inside triplets was violated so that, for example, a triplet-initial syllable was used only in foil-medial or foil-final position. While the first eight foils could only be rejected based on the detection of statistical violations, the latter eight foils could be rejected based on either (or both) statistical violations and positional memory mechanisms.

Triplets and foils were both used twice during the test, yielding 48 trials, and presented in a different random order to each participant. On each trial, participants first made a yes/no answer, which indicated whether they were able to endorse triplets and reject foils as constituents of the exposure stream. They then indicated their confidence in this decision on a 4-point scale. This revealed whether they assigned higher confidence to correct than incorrect responses, reflecting engagement of metacognitive processes.

Non-linguistic input, auditory modality

For non-linguistic stimuli, we used natural environmental sounds (water drops, footsteps, squeaks, animal noises, etc.) from https://freesound.org. These sounds were equalized in duration to 300ms and normalized in intensity to 80dB (to make them perceptually similar in terms of loudness). Twenty four sounds were arranged into triplets and 8 sounds were used as inter-triplet fillers, using a metrical structure identical to that used for the linguistic stimuli (see Figure YA); longer pauses within the metrical structure were set to 200ms and shorter pauses were set to 100ms. The sounds were concatenated into a continuous stream (each triplet was presented

80 times in the exposure stream). A linear intensity increase was implemented instead of the initial rising boundary tones used for the linguistic material, and a linear intensity decrease was used instead of falling boundary final tones. Amplitude ramping was applied over two initial and two final syllables on the edges of largest groups (to match IPs on linguistic material), and over one syllable on the edges of small groups within larger groups (to match PP-boundaries within IPs on linguistic material). These rampings were perceived as gradual fluctuations in loudness on the edges of larger constituents; fluctuations were not aligned with the triplet boundaries.

For the recognition test, we concatenated recurrent triplets of environmental sounds from the exposure input in isolation, with each sound lasting 300ms at normalized intensity, and 16 foils, made by concatenating sounds that never occurred consecutively during exposure (only those sounds which composed the recurrent triplets in the exposure were used for the foils). Also, half of the foils preserved the ordinal position of sounds inside the triplets, and half of the foils violated ordinal positioning. As in the recognition test on the linguistic material, each test item was presented twice, yielding 48 trials in total, and on each trial, we collected accuracy and confidence judgments.

Linguistic input, visual modality

Twelve syllables were selected and grouped into four triplets (TE-GU-BA, TA-BO-FA, KA-BE-TO, GA-FO-BU), and 8 syllables were used for inter-triplet fillers (TU, GO, GE, KO, KU, FU, FE, KE). Note that these syllables were different from those used in the auditory modality. The syllables were presented in the middle of the screen one by one, for 500ms each. TPs between syllables within triplets were 100%, while TPs between filler syllables and triplet-boundary syllables were 12.5%, matching those implemented in the auditory stream. We used the same frame to arrange fillers and triplets as in the auditory modality. We used punctuation marks (commas and periods, also presented on screen for 500ms) as the boundary cues between larger constituents (equivalent to PP clauses within IP sentences, for a stream of complex sentences consisting of two clauses each). In the exposure stream, each triplet was presented 50 times.

The recognition test included 4 triplets and 8 foils composed of the same syllables as the triplets, which had never occurred consecutively in the exposure stream. Syllables in half of the foils preserved the ordinal position of the triplets; in the other half of the foils, ordinal positioning within triplets was violated. Each item was presented twice (each syllable was displayed 500ms in the middle of the screen), yielding 24 trials. We collected accuracy and confidence judgments on each trial.

Non-linguistic input, visual modality

We generated 20 fractals at http://sirxemic.github.io/ifs-animator/, using 12 of these for four triplets and a further eight for inter-triplet fillers (Figure 3), then arranged them according to the same pattern as the elements in the other experiments (Figure 2a, with fractals instead of syllables). For the boundary signals (instead of punctuation marks in linguistic material), we used white squares between the largest structural units (aka the sentences in the linguistic material) and grey squares between the smaller structural units embedded within these larger structural units (aka the clauses within sentences in the linguistic material). The fractals were shown on a grey background, such that grey squares, which were only one tone lighter than the background, subjectively appeared less prominent than white squares. Each fractal or boundary cue was displayed in the center of the screen for 500ms. Each triplet appeared 50 times in the familiarization stream. Construction of items for the recognition test was similar to that for the linguistic material, but used fractals instead of syllables. During the recognition test, accuracy and confidence judgments on each trial were elicited.

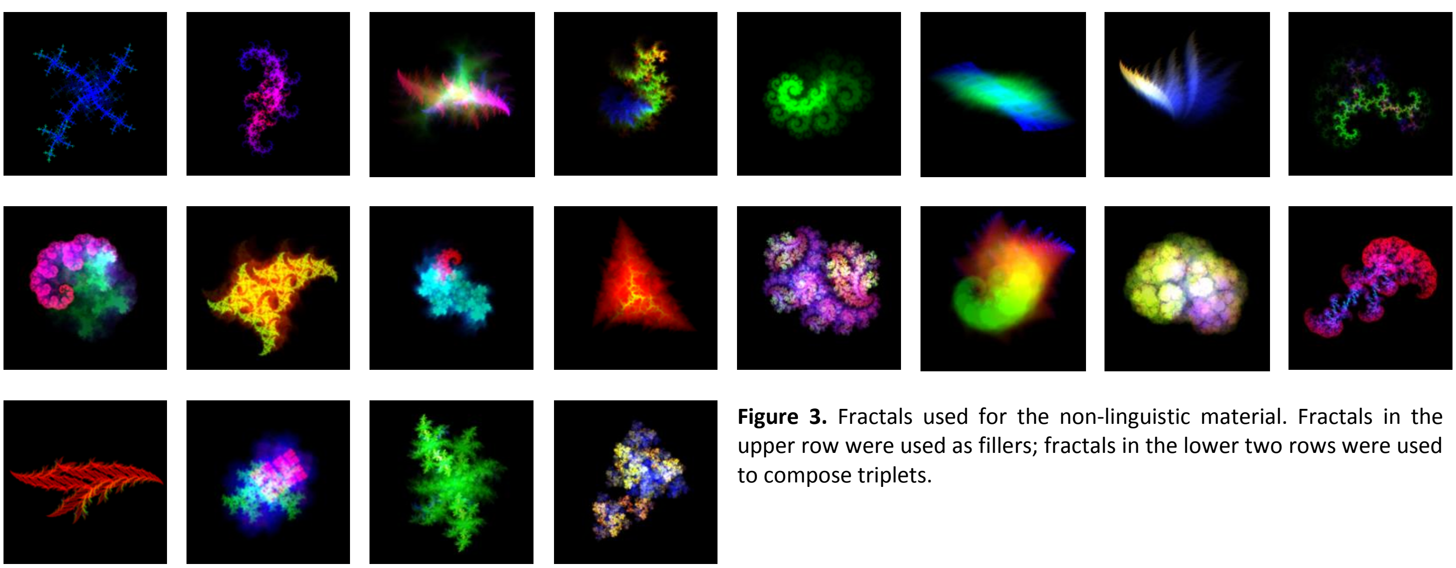

Figure 3. Fractals used for the non-linguistic material. Fractals in the upper row were used as fillers; fractals in the lower two rows were used to compose triplets.

Methodological approach to measuring metacognitive efficiency

Metacognitive control processes allow individuals to evaluate uncertainty, estimate the consequences of future actions and thus control behavior and decision-making. The second component of metacognition is metacognitive monitoring, which is the ability to evaluate one's own cognitive states and decisions that have already been made (Shimamura, 2008; Flavell, 1979; Dunlosky et al., 2007; Nelson & Narens, 1990). In the current work, we focus only on the latter. Evaluation involves estimating the likelihood of having made an error once a cognitive state has changed upon having made a decision (Kepecs et al., 2008). If the likelihood of making an error is high, individuals assign low confidence ratings to such decisions. Thus, a correct estimation of the likelihood of an error results in confidence ratings that discriminate between correct and incorrect responses. The higher the difference between confidence ratings assigned to responses that turned out to be correct than incorrect, the more efficient the metacognitive monitoring (Kunimoto et al., 2001; Persaud et al., 2007; Schwiedrzik et al., 2011).

Here, we use the signal detection theoretic approach (SDT) to measure metacognitive efficiency (Barrett et al., 2013; Galvin et al., 2003; Maniscalco & Lau, 2012). Endorsements or rejections of triplets and foils during the recognition test are referred to as type-1 responses: endorsed triplets are *hits*, endorsed foils are *false alarms*, rejected triplets are *misses*, and rejected foils are *correct rejections*. Confidence ratings are referred to as type-2 responses, or metacognitive evaluations: high confidence on correct responses (type-1 hits and correct rejections) constitute *meta-hits*; high confidence on wrong responses (type-1 false alarms and misses) *meta-false alarms*; low confidence on correct responses constitute *meta-misses*, and low confidence on wrong responses *meta-correct rejections*. Taking this logic further, when confidence ratings are arranged on a greater than binary scale (in our experiment, we used a 4-point scale), it is possible to apply type-2 ROC analysis by estimating a pseudo-d' for each participant that optimally fits the type-2 confidence ratings rather than type-1 responses (Barret et al., 2013; Maniscalco & Lau, 2012; 2014; Galvin et al., 2003; etc.). This requires parameter estimation of type-2 sensitivity as predicted by a type-1 model derived by fitting a type-1 model to the observed type-2 data, either by using log-maximum–likelihood estimation (Barrett et al., 2013; Maniscalco & Lau, 2012) or Bayesian modeling (Fleming & Daw, 2017). This pseudo-d', referred to as *meta-d'*, reflects the reliability of each participant's confidence ratings in discriminating correct and incorrect responses, i.e., whether correct type-1 responses are consistently assigned higher confidence ratings than incorrect type-1 responses. Meta-d' thus reflects the metacognitive sensitivity of each individual. Importantly, it avoids *metacognitive bias*, that is, any individual tendency to assign higher or lower confidence to decisions. For example, one participant may be overconfident and tend to assign confidence rating "4" to correct and "3" to incorrect type-1 responses; another participant may be under-confident, and assign "2" to correct and "1" to incorrect type-1 responses. Nevertheless, the confidence ratings of both participants discriminate type-1 accuracy equally well; these participants have equal metacognitive sensitivity, and their metacognitive bias will not influence meta-d'. Similarly, meta-d' will not differ between two individuals, one of whom consistently assigns "1" to incorrect and "4" to correct decisions, while the other assigns "2" to incorrect and "3" to correct decisions. If they are equally consistent, their metacognitive sensitivity will also be similar. However, the first of these individuals has more efficient metacognitive processing, which is not always captured by comparing meta-d'. Moreover, meta-sensitivity can depend on the amount of type-1 information available for metacognitive processing, and thus scales with type-1 performance (higher type-1 accuracy, or d', leads to higher meta-d' if the type-1 decisions are made consciously and individuals are aware of what they have learnt). Therefore, it is difficult to compare individuals who operate on wider or narrower subscales of confidence ratings (e.g., between 1 and 4 versus between 2 and 3 to express confidence between decisions with a different likelihood of error), or to compare individuals with different performance levels in type-1 tasks. Additionally, some tasks may be intrinsically more difficult than others, resulting in lower type-1 performance. Therefore, if one needs to compare metacognitive skills at the group level across tasks with different levels of difficulty, it is useful to estimate metacognitive efficiency relative to type-1 performance. This is easily done by calculating the ratio between meta-d' and d', the *M-ratio*, referred to as metacognitive efficiency (Maniscalco & Lau, 2012; Fleming & Lau, 2014). The M-ratio is particularly useful when the goal is to compare metacognition in tasks that are served by different cognitive and neural processes, when comparing populations with intrinsically different cognitive levels of development (e.g., children and adults or patients and healthy individuals), and when type-1 performance is measured in different units and on different scales (because meta-d' and d' are measured on the same scale, the ratio is dimensionless and comparisons can be made even though the scales and units of type-1 measurement differ).

We are interested in metacognitive processing in statistical learning across domains and modalities. Statistical learning is modality-specific (Conway & Christiansen, 2005; Frost et al., 2015) and better tuned to extract temporal regularities in the auditory modality and spatial patterns in the visual modality (Conway & Christiansen, 2009) but both our auditory and visual tasks require segmentation of recurrent constituents in the temporal dimension. To accommodate this difference, the number of triplets was smaller in the visual modality. Different low-level perceptual mechanisms serve statistical learning in different modalities (Conway, 2020; Siegelman et al., 2017). Fundamental differences between modalities and individual-specific differences make direct cross-modality comparison of type-1 performance in statistical learning almost impossible. But, by using SDT and estimating *M-ratios*, which take individual type-1 performance into account, we can directly compare metacognitive efficiency across domains and modalities in statistical learning. We are no longer limited by these fundamental differences in the perceptual and cognitive mechanisms applied or by type-1 performance across tasks.

For the purposes of this study, we adopted a method based on hierarchical Bayesian estimation of meta-d' (Fleming, 2017) (code available at https://github.com/metacoglab/Hmeta-d). Unlike maximum-likelihood estimation methods, which have their own advantages (not relevant for our particular study), Bayesian estimation naturally accounts for situations where the number of trials per confidence level differs within individuals and thus does not require “data padding” when a particular participant gives zero responses with a particular confidence level. This decreases sampling subjectivity by reducing the effect of a single outlier on group results and allows all data to be retained without removing outliers. It is robust when comparing groups with different number of participants and tasks with different number of trials.

Results:

Main analysis

Table 1 provides the M-ratio scores (i.e., measures of metacognitive efficiency) for each group (Basque-Spanish bilinguals, Catalan-Spanish bilinguals, and Spanish monolinguals), modality (Visual, Auditory) and domain (Linguistic, Non-linguistic). Estimated measures of metacognitive efficiency were compared using *domain* and *modality* as within-subject factors and *group* as a between-subject factor (Figure 4).

Table 1. Mean values and standard deviations (in brackets) of means for M-ratio scores in different domains and modalities in three groups.

	Auditory		**Visual**	
	Linguistic	**Non-linguistic**	**Linguistic**	**Non-linguistic**
Basque-Spanish bilinguals (N=46)	.696 (.069)	.643 (.194)	.451 (.011)	.712 (.182)
Catalan-Spanish bilinguals (N=43)	.522 (.091)	.689 (.055)	.277 (.015)	.702 (.322)
Spanish monolinguals (N=39)	.39 (.012)	.702 (.038)	.47 (.079)	.714 (.14)

An omnibus ANOVA showed a significant effect of all factors at $p<.0005$. Interpretation of these effects is, however, difficult due to strong and significant 3-way interaction between all factors, $F(2,125)=21.491$, $\lambda=.744$, $p<.0005$. All two-way interactions were also significant and strong: *domain*group*, $F(2,125)=30.895$, $\lambda=.669$, $p<.0005$; *modality*group*, $F(2,125)=16.608$, $\lambda=.79$, $p<.0005$; and *domain*modality*, $F(2,125)=44.087$, $\lambda=.739$, $p<.0005$.

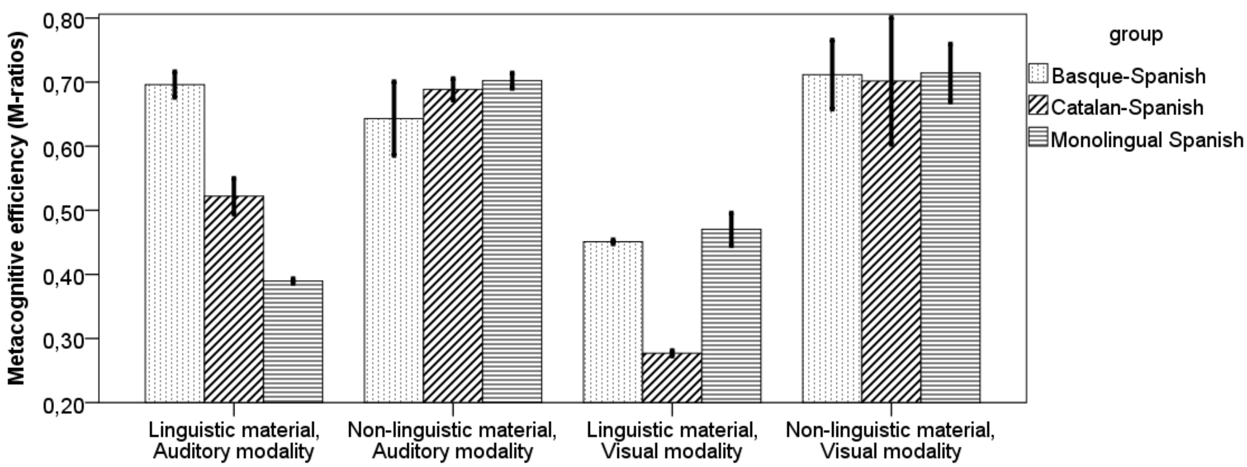

Figure 4. Metacognitive efficiency across groups, modalities, and domains. Error bars stand for 95% CI.

To tease apart these complex interactions, we compared M-ratios in the auditory modality in the linguistic domain. A Welch ANOVA showed that M-ratio means differed significantly between groups, $F(2,61)=481.799$, $p<.0005$. All pairwise comparisons remained significant at $p<.0005$ after applying Bonferroni correction, with the highest values of metacognitive efficiency in the Basque-Spanish bilingual group and the lowest values in the Spanish monolingual group. Catalan-Spanish bilinguals showed intermediate M-ratio scores. A comparison of M-ratio scores in the auditory modality in the non-linguistic domain did not reveal significant differences between group M-ratio means, $F(2,76)=2.6$, $p=.081$.

Additionally, we used paired t-tests (Bonferroni-corrected) to compare the M-ratios in the auditory modality within groups between domains. Metacognitive efficiency of the Basque-Spanish did not differ significantly, $t(45)=1.803$, $p=.234$, across the linguistic and non-linguistic domains. However, the meta-efficiency of Catalan-Spanish bilinguals, $t(42)=10.589$, $p<.0005$, and Spanish monolinguals, $t(38)=48.43$, $p<.0005$, was higher on non-linguistic than on linguistic material. However, meta-efficiency was not correlated at an individual level between domains for any group ($r=.098$, $p=517$ for Basque bilinguals, $r=.072$, $p=.649$ for Catalan bilinguals, and $r=.075$, $p=.649$ for Spanish monolinguals). This importantly suggests that the Basque-Spanish did not benefit from a transfer of metacognitive efficiency between domains in the auditory modality.

We next used a Welch ANOVA to compare M-ratios in the visual modality in the linguistic domain, $F(2,69)=1951.62$, $p<.0005$. Pairwise comparisons (with Bonferroni-corrected p-values) showed that metacognitive efficiency was lower in the Catalan-Spanish than in either the Basque-Spanish bilingual ($p<.0005$) or Spanish monolingual ($p<.0005$) groups. We did not observe any difference in meta-efficiency between Spanish monolinguals and Basque-Spanish bilinguals, $p=.15$. Comparison of M-ratio scores in the visual modality in the non-linguistic domain did not reveal significant differences in M-ratio means between groups, $F(2,79)=.026$, $p=.974$.

Next, we used paired t-tests to compare M-ratio scores between domains within the visual modality. The results showed that in the visual modality, meta-efficiency was higher in the non-linguistic than linguistic domain, in all groups: Basque-Spanish bilinguals, $t(45)=9.808$, $p<.0005$; Catalan-Spanish bilinguals, $t(42)=8.652$, $p<.0005$; Spanish monolinguals, $t(38)=9.334$, $p<.0005$. Absence of significant correlations at the individual level between M-ratios on linguistic and non-linguistic material in the Basque bilingual ($r=.191$, $p=.203$), Catalan bilingual ($r=.022$, $p=.888$) and Spanish monolingual ($r=.031$, $p=.85$) groups suggests there is no association (or transfer) between metacognitive efficiency across different domains.

Supplementary analysis

As supplementary analyses, we first estimated M-ratios by considering only those foils where the positional order of sub-units (syllables, fractals, or sounds) was preserved (sub-units presented in the triplet-initial, medial, or final position during exposure had the same position in the foil). Then, we estimated M-ratios by modeling the noise represented only by those foils for which the positional order of sub-units was not preserved.

Foils that preserved sub-unit positions

A Welch ANOVA showed that M-ratios differed significantly between groups in the auditory modality in the linguistic domain, F(2,75)=4534.534, p<.0005, in the auditory modality in the non-linguistic domain, F(2,73)=15.963, p<.0005, in the visual modality in the linguistic domain, F(2,71)=244, 189, and in the visual modality in the non-linguistic domain, F(2,81)=6.906, p=.002. Figure 5 shows this pattern.

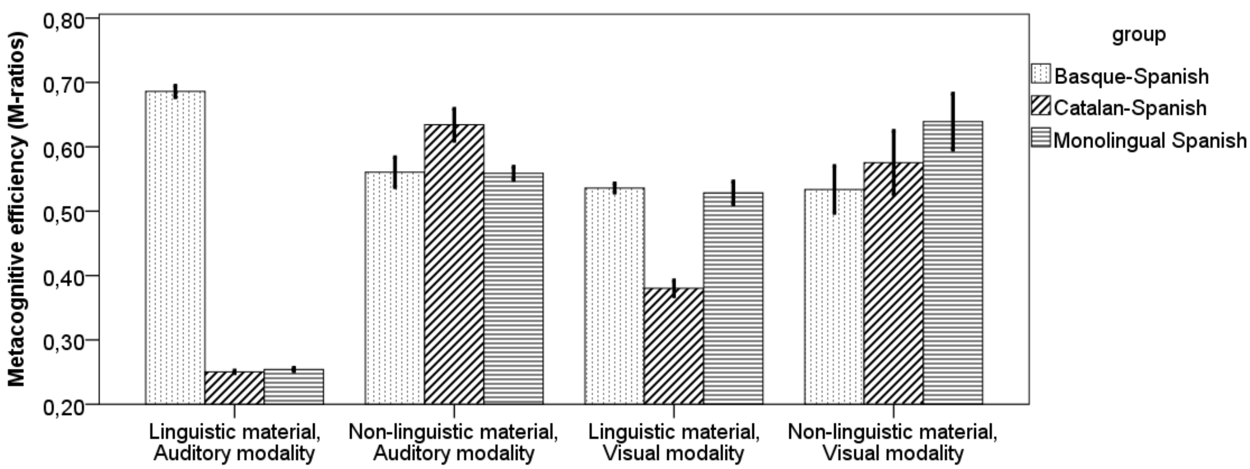

Figure 5. Metacognitive efficiency across groups, modalities, and domains; noise is represented only by foils, in which the relative position of sub-units (syllables, fractals or sounds) was preserved. Error bars stand for 95% CI.

Pairwise comparisons (all p-values Bonferroni-corrected) showed that in the auditory modality in the linguistic domain, metacognitive efficiency was significantly higher in the group of Basque-Spanish bilinguals (p<.0005 for both pairwise comparisons), while no difference was observed between Catalan-Spanish bilinguals and Spanish monolinguals (p=1.0). Intriguingly, in the non-linguistic domain, metacognition was higher in the group of Catalan-Spanish bilinguals than in the group of Basque-Spanish bilinguals (p<.0005) or Spanish monolinguals (p<.0005), while no significant difference in metacognitive efficiency in the auditory modality in the non-linguistic domain was observed between Basque-Spanish bilinguals and Spanish monolinguals. In the visual modality, on linguistic stimuli, meta-efficiency was significantly lower in Catalan-Spanish bilinguals than in the other two groups (at p<.0005), with no difference between Spanish monolinguals and Basque-Spanish bilinguals (p=1.0). Finally, in the visual modality in the non-linguistic domain, we did not observe significant differences (after correcting p-values) between the bilingual groups.

Foils that violated sub-unit positions

A Welch ANOVA showed that M-ratios between groups differed significantly for the auditory modality in the linguistic domain, F(2,60)=167.57, p<.0005, in the auditory modality in the non-linguistic domain, F(2,77)=331.214, p<.0005, in the visual modality in the linguistic domain, F(2,58)=1071.463, and in the visual modality in the non-linguistic domain, F(2,82)=24.742, p<.0005. Figure 6 shows this pattern.

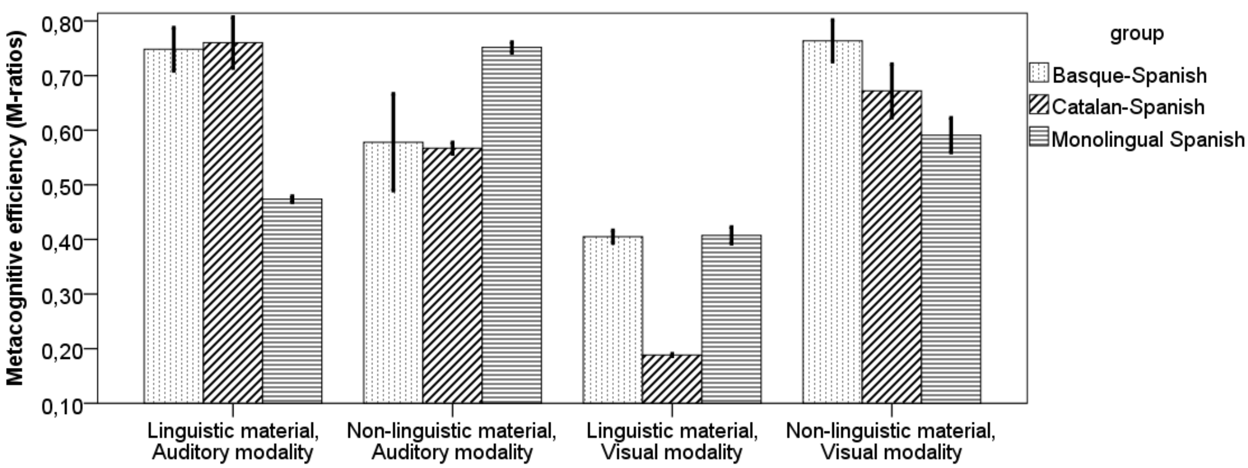

Figure 6. Metacognitive efficiency across groups, modalities, and domains; noise is represented only by foils, in which the relative position of sub-units (syllables, fractals, or sounds) was violated. Error bars stand for 95% CI.

Pairwise comparisons showed that in the auditory modality in the linguistic domain, metacognitive efficiency was lowest in the group of monolinguals (p<.0005), with no significant difference in metacognitive efficiency between the two bilingual groups (p=1). In the non-linguistic domain, however, M-ratio scores were higher in the monolingual than the bilingual groups (p<.0005), with no significant difference between bilingual groups (p=1). In the visual modality in the linguistic domain, we observed significantly lower M-ratio scores in Catalan-Spanish bilinguals than in the other two groups (p<.0005), with no significant difference between Basque-Spanish bilinguals and Spanish monolinguals (p=1). Finally, in the visual modality in the non-linguistic domain, efficiency was higher in Basque-Spanish bilinguals than in Catalan-Spanish bilinguals (p=.004), and in the latter group, efficiency was higher than in the monolingual group (p=.018).

Discussion:

Our analysis shows that overall, meta-efficiency is higher on non-linguistic than linguistic material, across both tested modalities. The hypothesis that bilingualism with typologically different languages might confer a metacognitive advantage was confirmed only in the auditory modality in the linguistic domain. Constant exposure to a more diverse range of language structures (via simultaneous acquisition and use of two typologically different languages) enhanced metacognitive efficiency in Basque-Spanish bilinguals. In the visual modality, Catalan-Spanish bilinguals exhibited lower metacognitive efficiency on linguistic material. Catalan is typologically similar to Spanish, so the linguistic structures Catalan-Spanish bilinguals need to acquire are similar in both languages, and do not require development of different processing strategies. We propose that an increase in the quantity of linguistic structures with no increase in the diversity of these structures may have a negative effect on meta-efficiency, resulting in lower metacognitive processing skills in the visual modality. Statistical learning is not a single mechanism, but rather an ability that is served by a set of neuro-cognitive mechanisms (Conway, C. & Christiansen, 2005; 2006; Frost et al., 2015; Polyanskaya, accepted to M&C; Thiessen, et al. 2013). The differential effects of linguistic experience on metacognitive monitoring across perceptual modalities in our task could be due to the need to monitor different sets of cognitive mechanisms in the visual and auditory modalities. An earlier study also reported that bilingual experience impeded meta-monitoring in a two-alternative-forced-choice task, in which participants saw two white circles on a black background on the screen and indicated whether the left or the right circle contained the most dots by pressing the appropriate arrow key on a standard computer keyboard (Folke et al., 2016). That result suggested that metacognitive efficiency in low-level perceptual tasks in the visual modality can indeed be inhibited by task-unrelated factors like linguistic experience; the mechanisms engaged by the dot discrimination task are also activated in statistical learning tasks (i.e., included in the set of cognitive mechanisms underlying statistical learning in the visual, but not in the auditory modality). Finally, Ordin et al. (2021) showed that linguistic experience has differential effects on statistical learning in the visual and auditory modalities. Written language, unlike spoken speech, is a relatively recent cultural innovation and statistical learning is more adapted for processing linguistic information in the auditory than in the visual modality. It is thus not surprising to observe different result patterns in different perceptual modalities or to find that individual differences in linguistic experience can impede metacognition in language tasks.

Ordin et al. (2020b) suggested that statistical learning mechanisms are sensitive to violations of statistical structure and thus detect deviations from extracted regularities (rather than compliance with this structure). We used two different types of foils, and they present different types of violations: foils that preserved syllabic positions but violated the learned transitional probabilities (TPs) between consecutive syllables; and foils that violated both TPs and the positional order of syllables (relative to constituent edges). Modeling noise caused by these different foil types supports our main finding and allows us to extend it further. Again, we found a consistent and convincing positive effect of bilingualism on metacognition – tracking cognitive performance in statistical learning – only in the auditory modality for linguistic material. When we included only the foils with position and TPs-violations in our metacognition models, we observed a general effect of bilingualism: both Catalan-Spanish and Basque-Spanish bilinguals were more efficient than monolinguals in tracking their own cognitive performance. But when we only included foils in which the position of sub-units was preserved, metacognitive efficiency was severely compromised in Catalan-Spanish bilinguals and Spanish monolinguals, but remained at the same level in Basque-Spanish bilinguals. This revealed that the efficiency of metacognitive monitoring in the statistical learning task was enhanced by simultaneous acquisition and use of typologically different languages. This pattern suggests that exposure to typologically different languages facilitates monitoring those specific cognitive processes that are engaged in TP-based segmentation of a continuous speech-like auditory input. Tracking performance of cognitive processes underlying positional memory in a sequence is overall easier than tracking TP-based computations. Both Basque-Spanish and Catalan-Spanish bilinguals are exposed to a larger volume of speech cues and linguistic structures by their simultaneous acquisition and use of two languages, but the structural diversity of the input is higher for the Basque-Spanish group. However, we did not see any difference in metacognitive efficiency between the bilinguals with typologically different and typologically similar languages when foils included the positional manipulation. Hence, we suggest that it is the volume of structures and cues rather than diversity of these structures and cues that increases the metacognitive efficiency of monitoring positional memory processes. By contrast, diversity is an important factor for monitoring TP-based statistical learning.

Statistical learning includes a set of mechanisms that are differentially engaged in different domains and modalities (Conway, 2020; Siegelman et al., 2017; Polyanskaya, in press; Thiessen et al., 2013), while metacognitive processes monitor the performance of these underlying cognitive mechanisms on a case-by-case (or mechanism-by-mechanism) basis. We did not observe any correlation between the metacognitive monitoring processes engaged by different modalities or domains. This suggests that monitoring a subset of auditory statistical language learning mechanisms might be enhanced, while monitoring a different subset of visual statistical learning mechanisms might be inhibited, in the same individual. Moreover, even within the same domain and modality (e.g., linguistic domain, auditory modality), statistical learning relies on a range of cognitive mechanisms (Frost et al., 2015), and the ability to track performance of these mechanisms is affected by external factors in a nuanced way (which was reflected in our study in different patterns of results when different types of foils were considered separately). In a nutshell, our hypothesis was confirmed only in the auditory modality: exposure to typologically different languages in a bilingual environment enhances metacognitive efficiency in a statistical learning task on linguistic material. In a more nuanced analysis, simultaneous acquisition and use of typologically different languages enhances metacognitive efficiency of TP-based learning, while simultaneous exposure to and acquisition of two languages, whether or not they are typologically different, enhances the efficiency of positional memory mechanisms during sequence learning tasks.

Another explanation for why the effect of typological difference was observed for foils that differed only in terms of transition probabilities but not relative positions could relate to cognitive load. It is harder to reject foils in which only TPs differ than to reject foils in which both TPs and the within-triplet position of syllables relative to triplet edges differ. Under higher cognitive load, the probability of making an error also increases, leading to higher activation of the error-detection mechanisms (De Bourbon-Teles, 2014; Ordin et al., 2020; Soto et al., 2019) that underlie metacognitive monitoring (Ordin et al., 2020). Thus, the bilingual advantage in metacognitive efficiency is stronger in the condition where people were likely more sensitive to the probability of committing an error, i.e., when the task became more difficult.

The reported study is the first in a project that aims to investigate the proximate mechanisms by which linguistic experience influences non-verbal behavior. The link between language and behavior has been documented (e.g., Levinson, 2003; Levinson & Jaisson, 2005), and it is believed that this influence is mediated by cognition. We know how language can influence cognitive processing and tune general cognitive mechanisms for processing the specific structures and cues of a particular ambient language or for handling linguistic code in general (Gleitman & Papafragou, 2005; Ordin et al., 2021; Costa & Sebastian-Galles, 2014). We also know that this effect can be transferred to non-language tasks (Marcus et al., 2007; Martin & Culbertson, 2020; Costa, Foukard, Arnon et al., 2014). However, the proximate mechanisms that translate cognitive changes induced by linguistic experience to non-verbal behaviour and decision-making in non-language domains have not been identified. I believe that this link is not direct but is mediated by metacognition, or, to be more exact, by the ability to monitor cognitive performance. Metacognitive enhancement in one domain can influence metacognition in a different domain (Carpenter et al., 2019; Mazancieux, 2020) and the neural circuits underpinning metacognition are known to be partially task- and domain-independent (McCurdy et al., 2013; Morales et al., 2018). As metacognition is related to decision-making and to guiding future behavior (Flavel, 1979; Kepecs et al., 2008; Schraw, 1998; Smith et al., 2003), metacognitive enhancement in one domain can affect decision-making strategies in a different domain. Bilinguals who speak typologically different languages need to monitor the multiple cognitive strategies they engage to efficiently process diverse language structures. This appears to enhance metacognition in the language domain. This benefit might then be transferred to other tasks and domains and lead to changes in decision making in non-verbal behavior. The current study is the first step in my project that aims to test this model.

Author Note

The authors declare no conflict of interest or activities, financial or otherwise, that might be seen as influencing the research.

Research in Context

The reported study is the first in a project that aims to investigate the proximate mechanisms by which linguistic experience influences non-verbal behavior. The link between language and behavior has been documented, and it is believed that this influence is mediated by cognition. We know how language can influence cognitive processing and tune general cognition for processing the specific structures and cues of an ambient language. However, proximate mechanisms that translate cognitive changes induced by linguistic experience to non-verbal behaviour and decisions have not been identified. We believe that this link is not direct but is mediated by metacognition, or, to be more exact, by the ability to monitor cognitive performance. Metacognitive enhancement in one domain can influence metacognition in a different domain and the neural circuits underpinning metacognition are known to be partially task- and domain-independent. As metacognition is related to decision-making and to guiding future behavior, metacognitive enhancement in one domain can affect decision-making strategies in a different domain. Bilinguals who speak typologically different languages need to monitor the multiple cognitive strategies they engage to efficiently process diverse language structures. This appears to enhance metacognition in the language domain. This benefit might then be transferred to other tasks and domains and lead to changes in decision making in non-verbal behavior. The current study aims to test this model by focusing on the theory that exposure to typologically different languages may modulate metacognitive monitoring in language tasks.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics 20(3)*, 242–275.

Andreose, A. & Lorenzo, R. (2013). Geography and distribution of the Romance languages in Europe. In M. Maiden, J. Smith, & A. Ledgeway (Eds.). *The Cambridge History of the Romance Languages, Volume II: Contexts*. (pp. 283–334). Cambridge: Cambridge University Press.

Arciuli, J., Torkildsen, JvK., Stevens, D.J., & Simpson, I.C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology 5*: 747.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition 106*, 1382–1407.

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods 18*, 535–552.

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition, 112(3)*, 494–500.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19(2)*, 290–303.

Birdsong D (2018) Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. Front. Psychol. 9:81

Birdsong, D. (2009). Uninterpretable features: psychology and plasticity in second language learnability. Second Lang. Res. 25, 235–243

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in stimulus–stimulus inhibition. *Bilingualism: Language and Cognition, 17(03)*, 610–629.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11(2)*, 282–298.

Carpenter, J., Sherman, M.T., Kievit, R.A., et al. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General 148(1)*, 51-64.

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

Conway, C. & Christiansen, M. (2009). Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology 21*, 561–580.

Conway, C.M. & Christiansen, M. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science 17*, 905–912.

Conway, C.M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews 112*, 279–299.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? Nature Reviews. Neuroscience 15(5), 336.

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). “Piensa” twice: On the foreign language effect in decision making. Cognition, 130(2), 236–254.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113(2)*, 135–149.

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning, 16(2)*, 69–97.

Dunlosky, J., Serra, M. J., & Baker, J. M. C. (2007). Metamemory. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.). *Handbook of applied cognition* (pp. 137–160). New York, NY: John Wiley & Sons

Dutoit, T., & Leich, H. (1993). MBR-PSOLA: Text-To-Speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication 13(3-4)*, 435–440.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (Eds.). Romanian. In *Ethnologue: Languages of the W*orld. Twenty-third edition (2020). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com (accessed 14/02/2021).

Erickson, L., Thiessen, E. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review 37*, 66–108.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition 9(2)*, 288–307.

Flavell, J.H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-development inquiry. *American Psychologist 34(10)*, 906–911.

Fleming, S. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness 1*: mix007.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience 8*, 443. doi: 10.3389/fnhum.2014.00443

Fleming, S., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review 124(1)*, 91–114.

Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition 150*, 119–132.

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

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 & Review 10*, 843–876.

Gebhart, A.L., Newport, E., & Aslin, R. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review 16*, 486–490.

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

Gollan, T. H., Weissberger, G. H., Runnqvist, E., et al., (2012). Self-ratings of spoken language dominance: a multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition* 15, 594–615.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hard, B.M., Recchia, G. & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General 140*, 586-604.

Johnson, J. Newport, E. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a Second Language. Cognitive Psychology 21, 60–99

Kahta, S., & Schiff, R. (2019). Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia 25(2)*, 142–157.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test (2nd ed.)*. Circle Pines, MN: American Guidance Service

Kepecs, A., Uchida, N., Zariwala, H., & Mainen, Z. (2008). Neural correlates, computation and behavioral impact of decision confidence. *Nature 455*, 227–231.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology 48(1)*, 171–184.

Kievit, R.A., Davis, S.W., Griffiths, J, et al. (2016). A watershed model of individual differences in fluid intelligence. *Neuropsychologia 91*,186-198.

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

Ladd, R. (2008). *Intonational Phonology* (2nd Edition). Cambridge: Cambridge University Press.

Levinson, S.(2003).Space in language and cognition: Explorations in cognitive diversity. Cambridge University Press.

Levinson, S., & Jaisson, P. (2005). Evolution and Culture. MIT Press.

Long, M. H. (1990). Maturational constraints on language development. Studies in Second Language Acquisition 12, 251–285

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

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d', response-specific meta-d', and the unequal variance SDT mode. In S. M. Fleming & C. D. Frith (Eds.). *The cognitive neuroscience of metacognition* (pp. 25–66). Berlin: Springer.

Marcus,G.,Fernandes,K.,&Johnson,S.(2007).Infant Rule Learning Facilitated by Speech.Psych. Science 18, 387–391

Martin, A., & Culbertson, J. (2020). Revisiting the Suffixing Preference: Native-Language Affixation Patterns Influence Perception of Sequences. Psychological Science 31, 1107-1116.

Mazancieux, A. (2020). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. Journal of Experimental Psychology: General, 149(9), 1788-1799.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. J. Neuroscience, 33(5), 1897–1906.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning 62*, 302–331.

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. Journal of Neuroscience, 38(14), 3534–3546.

Nelson, T. (1996). Consciousness and Metacognition. *American Psychologist 51*, 102-116.

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation 26*, 125–173.

Nespor, M. & Vogel, I. (2007). *Prosodic Phonology*. Berlin: De Gruyter.

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., & Samuel, A. (2021). An evolutionary account of intermodality differences in statistical learning. *Annals of the New York Academy of Sciences 1486(1)*,76-89.

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.

Ordin, M., Polyanskaya, L., & Soto, D. (2020b). Neural bases of learning and recognition of statistical regularities. Annals of the New York Academy of Sciences, 1467(1), 60-76. Do0.1111/nyas.14299

Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R.E., Cowey, A., & Lau, H. (2011). Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *NeuroImage 58*, 605–611.

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

Polyanskaya, L. (in press). Cognitive mechanisms of statistical learning and segmentation of continuous sensory input. *Memory and Cognition*.

Rademaker RL, Pearson J. (2012) Training Visual Imagery: Improvements of Metacognition, but not Imagery Strength. *Frontiers in Psychology 3*: 224 doi: 10.3389/fpsyg.2012.00224

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition 6*, 492–502.

Reber, A.S., Walkenfeld, F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition 17*, 888–896.

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language 47*, 172–196.

Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of Neurodevelopmental Disorders 10*, 35.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. *Science 274*, 1926-1928.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language 35*, 606–621.

Scattone, D., Raggio, D. J., & May, W. (2012). Brief report: Concurrent validity of the leiter-R and KBIT-2 scales of nonverbal intelligence for children with autism and language impairments. *Journal of Autism and Developmental Disorders 42*, 2486–2490.

Schiff, R., Sasson, A., Star, G., & Kahta, S. (2017). The role of feedback in implicit and explicit artificial grammar learning: a comparison between dyslexic and non-dyslexic adults. *Annals of Dyslexia 67*, 333–355.

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

Schwiedrzik, C. M., Singer, W., & Melloni, L. (2011). Subjective and objective learning effects dissociate in space and in time. *Proceedings of the National Academy of Sciences of the United States of America 108*, 4506–4511.

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). New York, NY: Psychology Press.

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass 14(3)*, e12365. doi:10.1111/lnc3.12365

Siegelman, N., Bogaerts, L., Christiansen, M., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B (Biological Sciences) 372*. doi: 10.1098/rstb.2016.0059

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences 13*, 389–396.

Smith, J.D., Shields, W.E., & Washburn, D.A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioural and Brain Sciences 26(3),* 317–339.

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

Turk-Browne, N.B., J.A. Jungé & B.J. Scholl. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General 134*, 552–564.

Woll, N. (2018). Investigating dimensions of metalinguistic awareness: what think-aloud protocols revealed about the cognitive processes involved in positive transfer from L2 to L3. Language Awareness, 27, 167-185.