Language and Cognition (2023), 15: 2, 314–336
doi:10.1017/langcog.2022.38

ARTICLE

Individual differences in visual word recognition: the role of epistemically unwarranted beliefs on affective processing and signal detection

Daniel Huete-Pérez* and Pilar Ferré

Universitat Rovira i Virgili, Department of Psychology, Research Center for Behavior Assessment (CRAMC), Tarragona, Spain
*Corresponding author. Email: daniel.huete@urv.cat

(Received 28 July 2022; Revised 15 November 2022; Accepted 17 November 2022)

Abstract
Previous studies have brought conflicting results regarding the effects of valence and arousal in visual word processing. Some authors have pointed to participants' individual differences as one of the possible explanations for these inconsistencies. The main aim of the present research was to examine whether participants' individual differences in the level of epistemically unwarranted beliefs (EUB) contribute to these conflicting results. Therefore, participants who varied in their level of paranormal, pseudoscientific and conspiracy beliefs (assessed by self-report measures) performed a lexical decision task (LDT) and a recognition memory task. Linear mixed-effects models over LDT response times revealed that the effects of words' emotional content (both valence and arousal) were modulated by the degree of individuals' EUB. In addition, signal detection theory analyses showed that in the recognition task (but not in the LDT) response bias became more liberal as individuals' EUB increased. These patterns of effects were not general to all EUB instances. The obtained results highlight the need to consider participants' individual differences in affective word processing and signal detection. In addition, this study reveals some basic psychological mechanisms that would underlie EUB, a fact that has both theoretical and applied relevance.

Keywords: visual word recognition; emotional processing; paranormal beliefs; pseudoscientific beliefs; conspiracy beliefs; lexical decision task; recognition memory task

1. Introduction

Psycholinguistic research has revealed the existence of multiple word properties that influence word processing (for an overview, see Adelman, 2012; Pexman, 2012; Yap & Balota, 2015). These properties can refer to different features of words, such as sublexical (e.g., bigram frequency), lexical (e.g., word frequency), semantic (e.g., concreteness/imageability) or affective (e.g., valence) features. The delimitation of these word properties is relevant not only for methodological reasons (i.e., important variables to guarantee an adequate experimental control), but also for theoretical reasons, since they have guided/constrained the development of psycholinguistic models and theories (Yap & Balota, 2015). For instance, frequent words in a given language are more easily processed than infrequent words (i.e., frequency effect; although this could be explained by the diversity of contexts in which the word appears rather than its frequency of occurrence per se; see Adelman, 2012). Psycholinguistic models must, therefore, explain how variability in frequency influences word processing. This logic extends to each of the word properties identified in the literature. Despite the huge progress made in modelling these items' properties, the same cannot be said for the individual differences related to the speakers themselves. More concretely, although there has been a lot of research on language-related differences between clinical and non-clinical populations (see Ball et al., 2008), general population subjects' individual differences have traditionally been much less explored, if not largely overlooked, in psycholinguistic theories and methods (Kidd et al., 2018). Several authors have argued for the need to overcome this tradition, since moving away from considering subjects' individual differences as error variance to, instead, exploring them as tentative systematic effects may help in advancing towards richer and more realistic psycholinguistic knowledge (Baayen et al., 2008; Kidd et al., 2018; Yu & Zellou, 2019). For example, Yap et al. (2012) suggested that some of the inconsistencies found in the literature could be explained by subjects' individual differences. In the last decade, researchers seem to have progressively considered this point, as interest in subjects' individual differences beyond clinical populations is growing in psycholinguistics (Kidd et al., 2018). For example, psycholinguistic studies have been conducted in the general population to explore subjects' individual differences such as language-related experience (see Yu & Zellou, 2019), vocabulary knowledge/size (e.g., Yap et al., 2012), age (e.g., Rossi & Diaz, 2016), executive functions (see Kidd et al., 2018; Yu & Zellou, 2019), affective-motivational particularities (see Fox, 2020), beliefs/opinions (see Fox, 2020) and autistic-like traits (see Yu & Zellou, 2019).

Within this context, some studies have shown that there are subjects' individual differences in relation to the processing of words' emotional content (e.g., Mueller & Kuchinke, 2016; Silva et al., 2012; Tárrega et al., 2021). Traditionally, the effects of affective word properties have been studied from a two-dimensional perspective, according to which the emotional content of words is basically described by two variables: valence [i.e., the extent to which the word is negative (e.g., *racism*), neutral (e.g., *poster*) or positive (e.g., *prize*)] and arousal [i.e., the degree of experienced activation in a range between deactivated-calmed (e.g., *librarian*) and activated-excited (e.g., *tornado*); see, e.g., Posner et al., 2005].[1] The effects of these two variables in word processing are still inconclusive. With respect to valence, while positive words are usually processed more easily than neutral words, negative words have produced mixed findings (advantage, disadvantage and null effects; see Hinojosa

[1] Although three dimensions were originally proposed (i.e., valence, arousal and dominance; see, e.g., Bradley & Lang, 1999), dominance explained much less variance and was less consistent than valence and arousal (Redondo et al., 2005), which probably has determined why the affective word processing literature has focused on valence and arousal (for an overview, see Hinojosa et al., 2020).

et al., 2020). Regarding arousal, mixed findings have been reported too: there are reports of facilitating (e.g., Recio et al., 2014), inhibitory (e.g., Kuperman et al., 2014) and null effects (e.g., Rodríguez-Ferreiro & Davies, 2019). Even though other word properties may contribute to these inconsistencies [e.g., concreteness (see Kousta et al., 2011; Borghi et al., 2017), semantic ambiguity (see Ferré et al., 2021) and word frequency (see Barriga-Paulino et al., 2022)], subjects' individual differences in relation to affective processing could also have a role in explaining these conflicting results (Mueller & Kuchinke, 2016; Silva et al., 2012). Just to give an example, in the study of Silva et al. (2012), participants with high- and low-disgust sensitivity performed a lexical decision task (LDT) which included both negative disgust-related words and neutral words. The effect of words' negative valence was modulated by participants' disgust sensitivity: an inhibitory effect for negative words in comparison to neutral words was found in the high-disgust sensitivity group, while this effect was facilitating in the low-disgust sensitivity group. Therefore, in a study in which there happened to be many disgust-related words, the valence effect in relation to negative words would vary depending on the group of participants: an inhibitory effect would arise if there are mostly high-disgust sensitivity participants, a facilitating effect would arise if there are mostly low-disgust sensitivity participants and even no significant effects are possible if there is a similar number of participants of each type in the sample (i.e., facilitating and inhibitory effects would cancel each other out). Therefore, like the case of disgust sensitivity, inconsistent results in valence and arousal effects across studies may be explained by the proportion/distribution of participants in relation to other individual differences which influence affective word processing. Following this rationale, any subjects' individual differences capable of provoking a systematic differential effect on the influence of words' emotional content is considered relevant.

In the present research, we will examine the role of individual differences in affective word processing, focusing on the level of *epistemically unwarranted beliefs* (EUB) of participants. EUB is a term used to refer to socially widespread claims that are not supported enough by either reliable empirical evidence or valid reasoning (Dyer & Hall, 2019), and it encompasses the paranormal (e.g., the existence of ghosts),[2] pseudoscience (e.g., complementary and alternative medicine of unproven efficacy) and conspiracy theories (e.g., Hitler did not die in 1945, but escaped and continued to live under a secret identity) (Lobato et al., 2014; Rizeq et al., 2020). These kinds of beliefs are not residual, but common in the general population (see Huete-Pérez et al., 2022). Therefore, study samples can easily vary in the distribution of these beliefs. In this context, we have chosen to explore this particular variable because of previous evidence suggesting differential sensitivity to affective word properties, at least in one of the three EUB dimensions. Concretely, Gianotti (2003), in her doctoral dissertation, observed that believers in the paranormal rated both positive and negative words as more extreme in valence than non-believers, hypothesising that paranormal believers would be more strongly influenced by both positive and negative emotional information. If the hypothesis of Gianotti (2003) is right, one

[2]The term 'paranormal' is not a unitary construct, but an umbrella term to groups topics such as afterlife (e.g., ghosts and reincarnation), extraordinary creatures (e.g., aliens and zombies), magic and mental powers (e.g., precognition and spells), mysticism (e.g., connection with the universe), religion (e.g., god/s and demons) and superstitions (e.g., number 13 bringing bad luck). See Irwin (2009) for an introduction to the concept and domains of paranormal beliefs.

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

would expect the effect of words' valence on word processing to be modulated in function of subjects' paranormal belief. Furthermore, if we accept that paranormal, pseudoscientific and conspiracy beliefs are instances of a broader category (i.e., EUB; Lobato et al., 2014; Rizeq et al., 2020) and, therefore, that they may share some characteristics and underlying mechanisms, words' valence effects might be expected to be modulated not only by the level of paranormal belief, but also by the levels of pseudoscientific and conspiracy beliefs. The main aim of the present study was to test, for the first time, this prediction. To this end, participants who varied in the degree of paranormal, pseudoscientific and conspiracy endorsement (as assessed by self-report measures) performed an LDT with words of the whole spectrum of valence and arousal values. The LDT was chosen because it is probably the most common experimental paradigm used to study the visual word processing of single words. In each trial of this task, participants are presented with a string of letters, and they then have to decide whether it is a real word in a particular language or something resembling a word, but that does not, in fact, exist in that particular language (i.e., a pseudo-word; Katz et al., 2012). Starting from the results and the hypothesis of Gianotti (2003), we expected to find an interaction between words' valence and subjects' levels of EUB, with larger valence effects for EUB believers than for non-believers. Gianotti (2003) focused on valence. However, as explained above, the emotional content of words has traditionally been defined not only in terms of valence, but also in terms of arousal. Consequently, we decided to explore also whether arousal effects in word processing are modulated by the degree of EUB (although we did not have a specific prediction here about the direction of the interaction).

Apart from exploring the interactive effects between words' emotional content and subjects' EUB, a secondary aim of the present research was to analyse participants' response patterns. Several studies have reported that paranormal believers tend to present a liberal response bias (also termed 'type I error bias'; see Brugger & Graves, 1997), that is, a bias towards making positive identifications of a target stimulus type irrespective of whether it is really present or not (e.g., Harrison et al., 2021; Krummenacher et al., 2010; Riekki et al., 2013; Rodríguez-Ferreiro & Barberia, 2021a). For instance, Riekki et al. (2013) presented inanimate pictures (objects, buildings, landscapes etc.) that contained face-like areas or not, and participants had to decide whether they saw any face in each picture. Despite the absence of significant differences in the ability to adequately discriminate between pictures that contained faces from those that did not contain them, both paranormal and religious believers showed a bias towards identifying faces in pictures in more cases than non-believers. Within the rationale of EUB being a grouping category for paranormal, pseudoscientific and conspiracy beliefs, we could expect this liberal response bias to be observed in these three EUB instances (see, e.g., Rodríguez-Ferreiro & Barberia, 2021a). Given these precedents, we expected to replicate this liberal response bias in the LDT (i.e., EUB believers would show a greater tendency towards saying 'yes, it is a real word' irrespective of the stimulus type, i.e., a word or a pseudo-word). However, we were not sure if there would be enough variability to find this effect because of the low error rate typically observed in this task. Consequently, we introduced an additional task, which is more error-prone. Indeed, immediately after the LDT, participants performed a recognition memory task. In the test phase of a recognition task, participants are presented with real words in a particular language, and they have to decide whether those words were previously presented in the encoding task – in this case the LDT – (*old words*) or not (*new words*). A liberal response bias in this task refers to the tendency to produce a 'yes, it is an old word' irrespective of the stimulus type. Considering the above, we expected this bias to be larger in EUB believers than in non-believers.

In a nutshell, the main purpose of the present research was to explore whether the effect of words' affective content over LDT response times (RTs) systematically varies as a function of individual differences in the EUB levels of participants. We expected to find larger valence effects for EUB believers than for non-believers, but we had no specific predictions regarding arousal. As a secondary aim, we expected to replicate the liberal response bias previously observed for EUB believers. We expected to clearly find this bias in the recognition task, but we were unsure if we would also find it in the LDT. Finally, another secondary aim was to explore whether the two predicted effects (interactive effects with words' emotionality and main effects in subjects' response bias) can be generalised across different instances of EUB. The degree to which an effect is either common or specific will depend on whether the same pattern of effects is more or less often observed through the different EUB dimensions.

2. Method

2.1. Participants

A convenience-volunteer sample of 99 undergraduate Psychology students from the Universitat Rovira i Virgili (URV, Tarragona, Spain) participated in the study. All participants gave their informed written consent and received academic credits for their participation. The study protocol was approved by the *Comitè Ètic d'Investigació en Persones, Societat i Medi Ambient* of URV (reference: CEIPSA-2021-TD-0023), and it was in accordance with the Declaration of Helsinki. Two participants were removed from the study since a technical problem prevented them to complete all the tasks. Therefore, there were 97 valid participants (79 women and 18 men) aged between 19 and 42 years ($M = 20.89$, $SD = 2.90$).

2.2. Materials

Our starting point was a database containing 3,842 Spanish words for which there were published values available for all the lexico-semantic properties of interest (i.e., age of acquisition, concreteness, emotional arousal, emotional valence, familiarity, lexical frequency, lexical length, lexical neighbourhood, semantic ambiguity, sublexical frequency, word prevalence and lexical similarity between Spanish–Catalan translations).[3] Age of acquisition ratings were obtained from Alonso et al. (2015) and Hinojosa et al. (2016). Concreteness ratings were obtained from EsPal (Duchon et al., 2013), Guasch et al. (2016) and Hinojosa et al. (2016). Emotional arousal and valence ratings were obtained from Guasch et al. (2016), Hinojosa et al. (2016) and Stadthagen-González et al. (2017). Familiarity ratings were obtained from EsPal (Duchon et al., 2013), Guasch et al. (2016) and Hinojosa et al. (2016). Two different variables of lexical frequency were obtained from the

[3]Since our participants are mostly Spanish–Catalan bilinguals, it is necessary to control the degree of cognate status (i.e., lexical overlap; see Guasch et al., 2013) between their translations.

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

subtitles database of EsPal (Duchon et al., 2013): word frequency and contextual diversity. In relation to lexical length, the number of letters for each word was obtained from EsPal (Duchon et al., 2013). Regarding lexical neighbourhood, three different variables were obtained from the subtitles database of EsPal (Duchon et al., 2013): number of orthographic neighbours, number of higher frequency orthographic neighbours and mean Levenshtein distance of the 20 closest words. The lexical similarity between Spanish–Catalan translations was indexed through the normalised Levenshtein distance between the two words obtained from NIM (Guasch et al., 2013). Semantic ambiguity was indexed through objective measures, more concretely the number of senses of the *Diccionario de la Lengua Española* (Real Academia Española, 2014; http://dle.rae.es/). Two different variables of sublexical frequency were obtained from the subtitles database of EsPal (Duchon et al., 2013): bigram frequency and trigram frequency (all mean, token-absolute). Finally, natives-from-Spain word prevalence ratings were obtained from Aguasvivas et al. (2018). Some of these ratings were recovered through EmoFinder[4] (Fraga et al., 2018).

2.2.1. Lexical decision task

Due to the time constraints of the experimental session, it was not feasible for participants to perform the LDT with all the available words. Consequently, a representative sample of 300 words was randomly selected, taking care of not having words from the same word family (e.g., *viajar* and *viajero*). Two-sample independent Kolmogorov–Smirnov tests were performed to ensure that the distribution of the selected words in the different lexico-semantic properties did not significantly differ from the distribution observed in the original word pool (all $p \geq .277$ when compared to the initial 3,842 words, including themselves). In addition, a visual inspection of histograms was performed to further ensure the similarity of the distributions. Table 1 shows the descriptive statistics of the 300 selected words for the LDT.

In addition, 300 pseudo-words were created with Wuggy (Keuleers & Brysbaert, 2010) to have the same number of 'yes' and 'no' responses in the LDT. These pseudo-words were matched to target words in subsyllabic structure, length and transition frequencies. Pseudohomophones were avoided (i.e., strings of letters that are orthographically pseudo-words but that share the phonology with a real word) both considering Spanish (e.g., *elar*) and Catalan (e.g., *rabe*). Furthermore, since it is important to not have 'systematic differences between the words and the nonwords, other than the fact that the former belong to the language and the latter do not' (Keuleers & Brysbaert, 2010, p. 628), accents were added to some pseudo-words (e.g., *érfato*).

2.2.2. Recognition task

Sixty words were randomly selected from the 300 words seen in the LDT in order to act as old words in the recognition task. Two-sample independent Kolmogorov–Smirnov tests were performed to ensure that the distribution of the selected words in the different lexico-semantic properties did not differ significantly from the

[4]EmoFinder has the added value of having rescaled some variables that were not collected using the same scale across databases (e.g., familiarity ratings).

Table 1. Descriptive statistics of the lexico-semantic properties for the 300 words used in the LDT

	M	*SD*	Range (min–max)
WPrev	2.35	0.25	1.35–2.58
Log_Frq	1.03	0.66	0.00–2.94
Log_Cont_Divers	0.72	0.49	0.00–1.97
Abs_tok_MBOF	26,249.52	10,649.56	4,809.53–62,619.60
Abs_tok_MTOF	2,907.97	2,637.87	11.61–19,212.56
Num_letters	7.29	2.04	3–12
N	4.06	6.12	0–35
NHF	0.52	1.50	0–14
Lev_N	1.98	0.63	1.00–3.80
NLD	0.75	0.25	0.00–1.00
Fam	5.27	1.04	2.03–6.86
AoA	7.17	2.00	1.72–10.80
Conc	4.75	1.00	2.17–6.71
Val	5.26	1.57	1.30–8.50
Aro	5.33	1.13	2.28–8.45
Dict_Sen	5.38	5.29	1–43

Note. WPrev = word prevalence (in *z*-scores); Log_Frq = word frequency (in logarithmic scale); Log_Cont_Divers = word contextual diversity (in logarithmic scale); Abs_tok_MBOF = bigram frequency (mean, token-absolute); Abs_tok_MTOF = trigram frequency (mean, token-absolute); Num_letters = number of letters; N = orthographic neighbours; NHF = orthographic neighbours of higher frequency; Lev_N = mean Levenshtein distance of the 20 closest words; NLD = normalised Levenshtein distance between Spanish–Catalan translations; Fam = familiarity; AoA = age of acquisition; Conc = concreteness; Val = emotional valence; Aro = emotional arousal; Dict_Sen = dictionary senses.

distribution observed in the LDT word pool (all $p \geq .468$ when compared to the initial 300 words seen in the LDT, including themselves). Sixty words were selected from the 3,542 words of the initial set that had not been included in the LDT to act as new words in the recognition task. We did not include words from the same family as the ones seen in the LDT (e.g., *humanidad* and *humano*). The selection of these words was performed with Match (van Casteren & Davis, 2007) to ensure a pairwise matching with the 60 old words in all the lexico-semantic properties considered. Independent samples *t*-tests showed that the matching was successful (all $p \geq .290$). In addition, two-sample independent Kolmogorov–Smirnov tests were performed to ensure that the distribution of the new words in the different lexico-semantic properties did not significantly differ from the distribution observed in the old words (all $p \geq .375$). Table 2 shows the descriptive statistics for the old and new words included in the recognition task.

Finally, the distribution of grammatical categories was also equivalent between old words (50 nouns, 13 adjectives and 11 verbs) and new words (51 nouns, 12 adjectives and 10 verbs).[5]

2.3. Procedure

Participants performed the experimental tasks in groups of three as follows. First, they signed an informed written consent to participate in the study. Second, they

[5]It should be considered that some words can be grammatically ambiguous, and in these cases, the word has been counted one time for each possible grammatical category (e.g., the word *trato* has been counted once as noun and once as verb). This fact explains why the grammatical category count adds up to 73 in new words and 74 in old words while there are only 60 words in each condition.

Table 2. Descriptive statistics of the lexico-semantic properties for the 60 old words and 60 new words used in the recognition task

	Old words			New words		
	M	*SD*	Range (min–max)	*M*	*SD*	Range (min–max)
WPrev	2.32	0.25	1.51–2.58	2.29	0.24	1.52–2.58
Log_Frq	1.04	0.67	0.00–2.94	1.06	0.65	0.01–3.04
Log_Cont_Divers	0.74	0.50	0.00–1.97	0.73	0.48	0.01–1.98
Abs_tok_MBOF	25,456.51	9,524.68	8,242.43–44,808.21	24,645.53	9,529.85	8,624.52–46,693.97
Abs_tok_MTOF	2,814.23	2,254.69	226.90–8,893.44	2,755.37	1,974.29	103.00–8,774.34
Num_letters	7.28	2.07	4–12	7.38	2.00	3–12
N	4.03	5.95	0–30	3.82	5.92	0–30
NHF	0.45	1.06	0–7	0.45	1.06	0–7
Lev_N	1.96	0.64	1.00–3.70	2.07	0.65	1.00–3.65
NLD	0.76	0.26	0.00–1.00	0.76	0.26	0.00–1.00
Fam	5.37	0.94	2.97–6.86	5.34	0.87	2.54–6.75
AoA	7.17	1.91	3.28–10.78	7.16	1.95	2.72–10.88
Conc	4.72	0.98	2.17–6.53	4.83	0.98	3.19–6.50
Val	5.30	1.46	1.80–8.18	5.56	1.23	2.30–7.65
Aro	5.31	1.17	3.15–8.45	5.30	0.96	2.90–7.60
Dict_Sen	4.98	4.22	1–24	5.02	4.62	1–26

Note. WPrev = word prevalence (in *z*-scores); Log_Frq = word frequency (in logarithmic scale); Log_Cont_Divers = word contextual diversity (in logarithmic scale); Abs_tok_MBOF = bigram frequency (mean, token-absolute); Abs_tok_MTOF = trigram frequency (mean, token-absolute); Num_letters = number of letters; N = orthographic neighbours; NHF = orthographic neighbours of higher frequency; Lev_N = mean Levenshtein distance of the 20 closest words; NLD = normalised Levenshtein distance between Spanish–Catalan translations; Fam = familiarity; AoA = age of acquisition; Conc = concreteness; Val = emotional valence; Aro = emotional arousal; Dict_Sen = dictionary senses.

performed an LDT. Third, immediately after the LDT, they performed a recognition task. Finally, they filled out two questionnaires: the Popular Epistemically Unwarranted Beliefs Inventory (PEUBI; Huete-Pérez et al., 2022) and the Pseudoscientific Belief Scale, Revised Version (PSEUDO-R; Fasce et al., 2021). At the end of the experimental session, participants were debriefed regarding the nature of the study if they so wished.

2.3.1. Lexical decision task

Each trial began with a fixation point ('+') appearing in the middle of the screen for 500 ms. Then the stimulus (Arial font, size 11, lowercase) replaced the fixation point, and participants had to decide whether the string of letters was a Spanish word (pressing the 'yes' button with the index finger of the dominant hand) or not (pressing the 'no' button with the index finger of the nondominant hand). The trial finished when participants responded or the time limit of 2,000 ms had elapsed. No feedback was given. Trials were administered in a continuous running mode with an intertrial interval of 750 ms, with a break every 150 stimuli (participants continued the experiment by pushing a foot pedal). Participants carried out 14 practice trials before starting the experimental trials. Stimuli presentation and responses recording were done with DMDX (Forster & Forster, 2003).

2.3.2. Recognition task

Each trial began with a fixation point ('+') appearing in the middle of the screen for 500 ms. Then a Spanish word (Arial font, size 11, lowercase) replaced the fixation point, and participants had to decide whether it had been seen in the previous LDT (pressing the 'yes' button with the index finger of the dominant hand) or not (pressing the 'no' button with the index finger of the nondominant hand). The trial finished when participants responded or the time limit of 3,000 ms had elapsed. No feedback was given. Trials were administered in a continuous running mode with an intertrial interval of 750 ms, without breaks. Participants carried out eight practice trials before starting the experimental trials. Stimuli presentation and responses recording was done with DMDX (Forster & Forster, 2003).

2.3.3. Popular Epistemically Unwarranted Beliefs Inventory (PEUBI)

Developed by Huete-Pérez et al. (2022), this inventory assesses five correlated dimensions of EUB [Superstitions (PEUBI-S), Occultism and Pseudoscience (PEUBI-OP), Traditional Religion (PEUBI-TR), Extraordinary Life Forms (PEUBI-ELF) and Conspiracy Theories (PEUBI-CT)] through 36 items on a 5-point scale (*1 = Fully disagree*, *5 = Fully agree*). In the original study, PEUBI showed good psychometric properties in terms of reliability as internal consistency (estimates $\geq .85$), reliability as temporal stability (estimates $\geq .75$), convergent validity and divergent validity. Since the range of pseudoscientific beliefs is somewhat restricted in PEUBI (i.e., mainly pseudoscience related to occultism and New Age), it was deemed appropriate to add a broader questionnaire of pseudoscientific beliefs.

2.3.4. Pseudoscientific Belief Scale, Revised Version (PSEUDO-R)

Developed by Fasce et al. (2021), this scale assesses pseudoscientific beliefs in a single factor/dimension through 19 items on a 5-point scale (*1 = Strongly disagree*, *5 =*

Strongly agree). In the original study, PSEUDO-R showed good psychometric properties in terms of reliability as internal consistency ($\alpha = .90$) and convergent validity.

2.4. Data analysis

2.4.1. Epistemically unwarranted beliefs scores

PEUBI offers the possibility of using both factor scores and the sum of raw scores (Huete-Pérez et al., 2022). In this case, we chose to use the sum of raw scores, that is, we added the scores of all the items of each factor (reverse scored items: 2, 5, 8, 11, 12, 14, 16, 25, 27, 28, 30 and 33). PSEUDO-R total scores were obtained by adding the raw scores of all its items (reverse scored items: 6 and 15).

2.4.2. Lexical decision task: response times

LDT RTs were analysed in R (version 4.1.3) with linear mixed-effects models (LMEM; see Baayen et al., 2008; Singmann & Kellen, 2020; Winter, 2019) using the following packages/libraries: car (version 3.0.12; Fox et al., 2022), effects (version 4.2.1; Fox et al., 2022), lme4 (version 1.1.29; Bates et al., 2022), LMERConvenienceFunctions (version 3.0; Tremblay & Ransijn, 2020), lmerTest (version 3.1.3; Kuznetsova et al., 2020), MuMIn (version 1.46.0; Barton, 2022), readxl (version 1.4.0; Wickham et al., 2022) and sjPlot (version 2.8.10; Lüdecke et al., 2021). The initial dataset contained 58,200 RTs (97 participants × 600 stimuli). However, since half of the stimuli were pseudo-words, only 29,100 RTs corresponded to real Spanish words. From this dataset, we removed 2,557 observations (8.79% of the total) corresponding to RTs of participants who committed >25% of errors (two participants), RTs of items with >70% of errors (none), RTs of incorrect responses (including those trials that reached the time limit of 2,000 ms), RTs < 300 ms and RTs > |2.5| *SD* of each participant's mean. Finally, we also removed 791 observations (2.72% of the total) corresponding to RTs > |2.5| *SD* above the residual mean of an LMEM which included only by-subject and by-item random intercepts (see, e.g., Tremblay & Tucker, 2011). Therefore, 25,752 RTs were finally included in the analyses.

Following Winter (2019), instead of opting for predetermined LMEM structures (e.g., by-default minimal or maximal random effect structures, which of both present associated problems), the construction of the model was theoretically driven, that is, guided by both the knowledge of the studied phenomenon (i.e., visual word processing) and the purposes of the study (i.e., examine whether emotional word processing is modulated by subjects' EUB). In addition, the decision of which model to construct was "made in advance, […] before starting to investigate the data" (Winter, 2019, p. 244). Consequently, our base model was an LMEM with non-transformed RTs as the dependent variable, word properties,[6]

[6]Not all word properties previously presented in Section 2.2 were included in LMEM for multicollinearity reasons. More concretely, three pairs of predictors presented $r \geq |.70|$ altogether with at least one variance inflation factor >3 (Winter, 2019): (1) logarithmic word frequency/logarithmic contextual diversity, (2) number of letters/mean Levenshtein distance of the 20 closest words and (3) number of orthographic neighbours/number of higher-frequency orthographic neighbours. In those cases, the strategy adopted was to remove from the analyses the second term of each pair.

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

an EUB score (only one score is used at a time; see below), trial order (to account for practice/learning and fatigue effects; see Baayen et al., 2008) and preceding trial (see Baayen et al., 2008) as fixed effects, and by-subject and by-item random intercepts:

> RT ~ Word prevalence + Logarithmic word frequency + Logarithmic bigram frequency + Logarithmic trigram frequency + Number of letters + Number of orthographic neighbours + Normalised Levenshtein distance between Spanish–Catalan translations + Familiarity + Age of acquisition + Concreteness + Valence + Arousal + Dictionary senses + Trial order + Preceding correct/incorrect response + EUB + (1 | subject) +(1 | item).

Hereinafter this first model will be referred to as *simple effects only model* (SEOM). To assess whether the predicted interactions were significant (i.e., Valence × EUB or Arousal × EUB), we had to create another LMEM identical to SEOM but with the addition of the interactive term [hereafter this second model will be referred to as *interactive effects added model* (IEAM)]:

> RT ~ Word prevalence + Logarithmic word frequency + Logarithmic bigram frequency + Logarithmic trigram frequency + Number of letters + Number of orthographic neighbours + Normalised Levenshtein distance between Spanish–Catalan translations + Familiarity + Age of acquisition + Concreteness + Valence + Arousal + Dictionary senses + Trial order + Preceding correct/incorrect response + EUB + Valence:EUB or Arousal:EUB + (1 | subject) +(1 | item).

Then, these two models were compared using likelihood ratio tests. If the comparison was significant and IEAM had better fit indexes (i.e., lower logLike and AIC values), the addition of the interactive effect of interest was justified. Otherwise, the SEOM was selected. Finally, using an adaptation of the tables for reporting LMEM of Meteyard and Davies (2020), the next information was extracted and reported from the final selected model: proportion of variance explained by the model (R^2), variance of each random effect and parameters of the fixed effects [b coefficients and its 95% confidence interval (Wald method), standard error, t statistic and significance p-value (Satterthwaite's method)].

A total of 12 model comparisons were performed because we had six possible EUB scores (PEUBI-S, PEUBI-OP, PEUBI-TR, PEUBI-ELF, PEUBI-CT and PSEUDO-R) and two possible interactive effects of interest (Valence × EUB and Arousal × EUB).

2.4.3. Lexical decision task: signal detection theory parameters

To explore the response bias in the LDT, correct and incorrect responses were analysed under the signal detection theory framework (for an overview of this theory, see Stanislaw & Todorov, 1999; see also Diependaele et al., 2012 for a discussion of analysing LDT under signal detection theory) through the following steps. First, only responses that were performed after 300 ms and before the time limit was reached (2,000 ms) were considered. Second, each valid observation was categorised as one of the four possible response types according to signal detection theory: hit ('yes' to real words), false alarm ('yes' to pseudo-words), miss ('no' to real words) and correct rejection ('no' to pseudo-words). Third, hit and false alarm rates were calculated as follows:

Hit rate = Hits/(Hits + Misses)

False alarm rate = False alarms/(False alarms + Correct rejections)

However, since any extreme hit or false alarm rate (i.e., 0% or 100%) prevents the calculation of the following steps from being carried out (Stanislaw & Todorov, 1999), we replaced 0% rates with $0.5/n$ and 100% rates with $(n - 0.5)/n$, being n the number of valid signal or noise trials (Stanislaw & Todorov, 1999). Fourth, each hit and false alarm rate was transformed into its corresponding z-score. Fifth, we calculated the following signal detection theory parameters: *d'* (a discriminability measure) and *C* (a response criterion measure). They were calculated following Stanislaw and Todorov (1999):

$$d' = z_{\text{Hit rate}} - z_{\text{False alarm rate}}$$

$$C = -0.5 \times (z_{\text{Hit rate}} + z_{\text{False alarm rate}})$$

The correlations between these signal detection theory parameters and EUB scores were analysed in R (version 4.1.3). Participants who committed >25% of errors (two participants) were removed.

2.4.4. Recognition task: signal detection theory parameters

Analogously to the LDT, the exploration of the response bias in the recognition task was performed through analysing correct and incorrect responses under the signal detection theory framework (see Rotello, 2017). The calculation of signal detection theory parameters for each participant was performed using the same procedure as in the LDT. The only two differences were the time cut-offs for valid responses (in this task, responses between 300 and 2,999.99 ms were considered because there was a time limit of 3,000 ms) and the definition of the four possible response types of the signal detection theory (i.e., hits were 'yes' responses to old words, false alarms were 'yes' responses to new words, misses were 'no' responses to old words and correct rejections were 'no' responses to new words).

The correlations between signal detection theory parameters and EUB scores were analysed in R (version 4.1.3). Following the criteria of previous recognition memory studies (e.g., Cortese et al., 2010, 2015), participants that committed >40% of errors (10 participants) were removed.

2.5. Statistical power and sample size

In comparison to more classical tests in which power/sample size calculations are standardised and considered easy to compute, carrying out those same calculations in the case of LMEM is complex and the procedures are not well stablished/are still under development (Feng, 2016; Meteyard & Davies, 2020). Fortunately, some guidelines have been proposed, and are being adopted in psycholinguistics. As Meteyard and Davies (2020) summarise, Scherbaum and Ferreter (2009) recommended using at least 30–50 participants and 30–50 items (i.e., 900–2,500 observations), whereas Brysbaert and Stevens (2018) recommended using at least

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

40 participants and 40 items (i.e., 1,600 observations) in order to ensure 'a properly powered reaction time experiment with repeated measures'. In any case, as a general rule, Meteyard and Davies (2020) advice that we should try to have as many cases as possible in both participants and items (which should not be confounded with having a lot of participants and few items or vice versa since the variability within a unit of analysis matters). Following these rules of thumb, the final number of observations analysed in the LDT (i.e., 25,752; 97 participants × 300 words after the trimming procedure) would suggest that our study is powered enough.

3. Results

3.1. EUB scores

Descriptive statistics of EUB scores for the 95 final participants of the LDT are presented in Table 3. On the one hand, PEUBI-OP, PEUBI-CT and PSEUDO-R scores are fairly normally distributed with adequate variability. Despite a moderate positive skew, PEUBI-S still shows enough variability. On the other hand, PEUBI-TR and PEUBI-ELF show such a highly positive skew (i.e., most participants scoring low) that they may be problematic for inferential purposes (e.g., a decrease in statistical power by range restriction; Hallgren, 2018).

The correlation matrix between the EUB scores of the 95 final participants of the LDT is presented in Table 4. As can be seen, most EUB scores were significant and positively correlated with each other, with the notable exception of PEUBI-TR.

3.2. Lexical decision task: response times

As previously stated, there were 12 separate analyses resulting from the six possible EUB scores and the two possible interactive effects of interest. For extension limitations, only a qualitative summary of the results is provided here. A complete report of all the analyses can be found in the Supplementary Material.

The Valence × EUB interactive effect was significant in one case (the model with PSEUDO-R as EUB score) and non-significant in the remaining five cases (models with PEUBI-S, PEUBI-OP, PEUBI-TR, PEUBI-ELF and PEUBI-CT as EUB score). When this interactive effect was significant, there was a clear linear facilitating effect of valence (the higher the valence of the word, the faster the RT; therefore, positive words produced faster RTs than neutral words, whereas negative words produced

Table 3. Descriptive statistics of EUB scores for the 95 final participants of the LDT

	M	*SD*	Range (min–max)	Skewness
PEUBI-S	15.53	5.93	7–31	0.54
PEUBI-OP	29.26	7.85	14–47	0.15
PEUBI-TR	11.25	5.49	6–29	1.23
PEUBI-ELF	10.14	3.39	6–21	0.95
PEUBI-CT	18.31	3.64	9–27	−0.23
PSEUDO-R	55.73	6.58	40–69	−0.21

Note. PEUBI-S = superstitions; PEUBI-OP = occultism and pseudoscience; PEUBI-TR = traditional religion; PEUBI-ELF = extraordinary life forms; PEUBI-CT = conspiracy theories; PSEUDO-R = pseudoscience.

Table 4. Correlation matrix between EUB scores for the 95 final participants of the LDT

	1	2	3	4	5	6
1. PEUBI-S	–					
2. PEUBI-OP	.63***	–				
	[.49, .74]					
3. PEUBI-TR	.10	.15	–			
	[−.10, .30]	[−.06, .34]				
4. PEUBI-ELF	.38***	.48***	.07	–		
	[.20, .54]	[.31, .62]	[−.13, 27]			
5. PEUBI-CT	.30**	.30**	.09	.49***	–	
	[.10, .47]	[.11, .47]	[−.11, 29]	[.31, .63]		
6. PSEUDO-R	.36***	.50***	.11	.40***	.43***	–
	[.17, .52]	[.33, .64]	[−.09, .31]	[.21, .55]	[.25, .58]	

Note. Pearson correlation coefficient (95% CI in brackets). PEUBI-S = superstitions; PEUBI-OP = occultism and pseudo-science; PEUBI-TR = traditional religion; PEUBI-ELF = extraordinary life forms; PEUBI-CT = conspiracy theories; PSEUDO-R = pseudoscience.
**$p < .01$.
***$p < .001$.

slower RTs than neutral words) in participants with high EUB scores. In contrast, the effects of valence on RTs progressively disappeared as the degree of belief in EUB decreased (see Fig. 1).

The Arousal × EUB interactive effect was significant in three cases (models with PEUBI-S, PEUBI-OP and PSEUDO-R as EUB score) and non-significant in the remaining three cases (models with PEUBI-TR, PEUBI-ELF and PEUBI-CT as EUB score). When this interactive effect was significant, there was a clear arousal linear facilitating effect (the higher the arousal of the word, the faster the RT) in participants with low EUB scores. In contrast, the effects of arousal on RTs progressively disappeared as the degree of belief in EUB increased (see Figs. 2, 3 and 4).

Apart from the above interactive effects, which were the main interest in this study, some of the other psycholinguistic predictors also showed significant effects: word prevalence, word frequency, word familiarity and the degree of Spanish–Catalan cognate status exerted a facilitating effect (i.e., the higher the value of the examined variable, the faster the LDT response), whereas bigram frequency, length

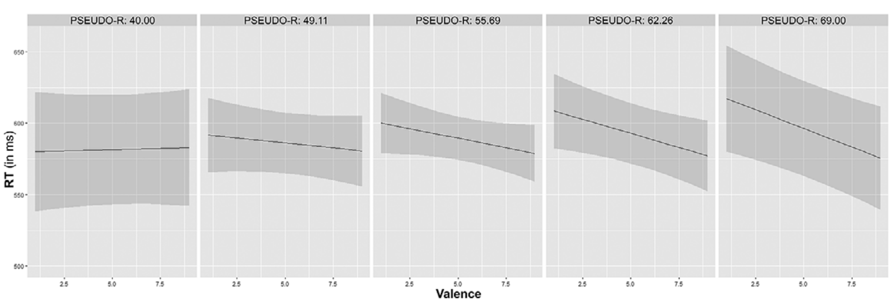

Fig. 1. Marginal effects of the interaction between valence and PSEUDO-R on LDT RTs. RT = response time; PSEUDO-R = pseudoscience. Each individual graph shows the effect of words' valence (ranging from 1 = *completely sad/negative* to 9 = *completely happy/positive*) over lexical decision task RTs in a particular representative value of the PSEUDO-R scores range. The grey band represents the 95% confidence interval.

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

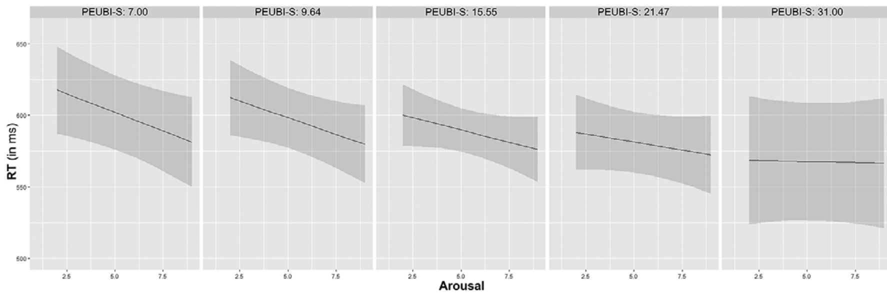

Fig. 2. Marginal effects of the interaction between arousal and PEUBI-S on LDT RTs. RT = response time; PEUBI-S = superstition. Each individual graph shows the effect of words' arousal (ranging from 1 = *completely quiet/calm* to 9 = *completely excited/energized*) over lexical decision task RTs in a particular representative value of the PEUBI-S scores range. The grey band represents the 95% confidence interval.

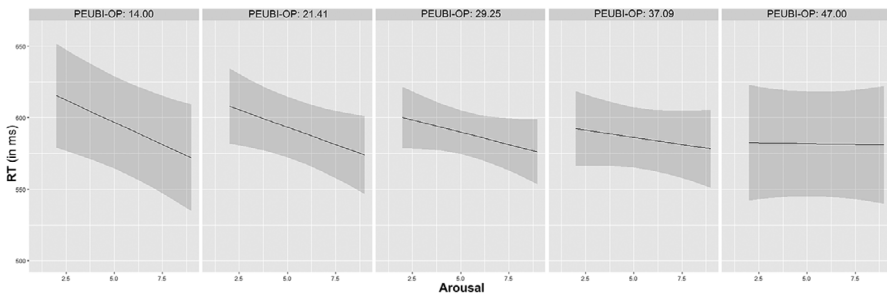

Fig. 3. Marginal effects of the interaction between arousal and PEUBI-OP on LDT RTs. RT = response time; PEUBI-OP = occultism and pseudoscience. Each individual graph shows the effect of words' arousal (ranging from 1 = *completely quiet/calm* to 9 = *completely excited/energized*) over lexical decision task RTs in a particular representative value of the PEUBI-OP scores range. The grey band represents the 95% confidence interval.

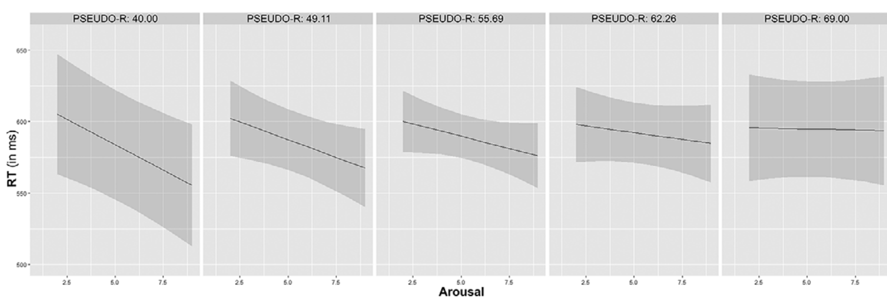

Fig. 4. Marginal effects of the interaction between arousal and PSEUDO-R on LDT RTs. RT = response time; PSEUDO-R = pseudoscience. Each individual graph shows the effect of words' arousal (ranging from 1 = *completely quiet/calm* to 9 = *completely excited/energized*) over lexical decision task RTs in a particular representative value of the PSEUDO-R scores range. The grey band represents the 95% confidence interval.

and age of acquisition exerted an inhibitory effect (i.e., the higher the value of the examined variable, the slower the LDT response). There were no significant effects of the following psycholinguistic variables: trigram frequency, number of orthographic neighbours, concreteness and number of dictionary senses. For an overview of the

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

effects of these word properties, see Adelman (2012), Pexman (2012) and Yap and Balota (2015).

Finally, there were also significant effects of trial order and preceding trial (see Baayen et al., 2008). More concretely, trial order exerted a facilitating effect (i.e., participants responded faster as they progressed through the task), and an error on the preceding trial exerted an inhibitory effect (i.e., making an incorrect response in the previous trial delayed the RT in a given trial).

3.3. Lexical decision task: signal detection theory parameters

The discriminability parameter *d'* was not significantly correlated with any EUB score (all $r \leq |.12|$, all $p \geq .232$), with the exception of PEUBI-TR, which was negatively correlated, although significantly marginally, $r(93) = -.20$, 95% CI [−.38, .00], $p = .053$. This means that, in general terms, the degree of EUB did not influence the ability to discriminate between real Spanish words and pseudo-words in the LDT.

The response criterion parameter *C* was not significantly correlated with any EUB score (all $r \leq |.13|$, all $p \geq .193$). This means that the degree of EUB did not influence the bias towards a 'yes' or 'no' response in the LDT.

3.4. Recognition task: signal detection theory parameters

The discriminability parameter *d'* was not significantly correlated with any EUB score (all $r \leq |.09|$, all $p \geq .408$), with the exception of PEUBI-CT, which was negatively correlated, although significantly marginally, $r(85) = -.20$, 95% CI [−.40, .00], $p = .054$. This means that, in general terms, the degree of EUB did not influence the ability to discriminate between old and new words in the recognition task.

The response criterion parameter *C* showed a significant negative correlation with PEUBI-ELF, $r(85) = -.24$, 95% CI [−.43, −.03], $p = .024$, PEUBI-CT, $r(85) = -.21$, 95% CI [−.40, −.00], $p = .048$ and PSEUDO-R, $r(85) = -.26$, 95% CI [−.45, −.05], $p = .015$, whereas it was not significantly correlated with PEUBI-S, PEUBI-OP and PEUBI-TR (all $r \leq |.15|$, all $p \geq .166$). The negative correlation observed between some instances of EUB and *C* means that the higher the level of EUB, the more liberal the response criterion (i.e., a bias towards saying 'yes').

4. Discussion

Given the inconsistencies in the literature regarding the effects of affective word processing, the main aim of the present research was to examine whether the effects of words' emotional content (i.e., valence and arousal) were, indeed, modulated by the degree of individuals' EUB. With this purpose in mind, participants who varied in the level of paranormal, pseudoscientific and conspiracy beliefs (as assessed by self-report measures) performed an LDT. The analyses evidenced that such modulation indeed exists: there were interactive effects between words' emotional content and participants' EUB over LDT RTs. A secondary aim was to explore if the liberal response bias observed for EUB believers in previous studies could be replicated in the present research. Signal detection theory analyses revealed that response bias became more liberal as individuals' EUB increased in the recognition task, but not in

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

the LDT. Finally, we intended to evaluate the extent to which the effects of interest of this study generalised across different EUB instances. Interactive effects with words' emotional content over LDT RTs occurred for pseudoscientific, occultist and superstitious beliefs (but not for religious, extraordinary creatures-related or conspiracy beliefs), and main effects over signal detection theory response criterion occurred for pseudoscientific, creatures-related and conspiracy beliefs (but not for superstitious, occultist or religious beliefs).

The modulation of affective word processing by EUB found here is not entirely surprising, considering prior evidence of affective differences in function of the level of EUB (e.g., paranormal, pseudoscientific and conspiracy beliefs have been linked to negative emotional states; see Douglas et al., 2020; French & Stone, 2014, Chapter 3; Galasová, 2022). In fact, previous studies had already found differences between believers and non-believers in the affective rating of emotional words (Gianotti, 2003). However, to the best of the authors' knowledge, this is the first study demonstrating that individual differences in EUB can make the effects of words' valence and arousal appear or disappear in on-line measures (i.e., RTs). The distinction between on-line and off-line measures (see Veldhuis & Kurvers, 2012) is crucial to understand our contribution here: whereas off-line measures/tasks involve responses that may be consciously influenced to a greater or lesser extent (e.g., when participants rate the valence of a word without any time limit, it is possible to respond in function of the expected social value instead of the subjective experienced value), on-line measures/tasks leave almost no room to the effect of consciously controlled processes (e.g., because of the time pressure of the task, such as in LDT). Therefore, on-line measures/tasks are more likely to reflect the automatic underlying cognitive processes of the task than off-line measures (Veldhuis & Kurvers, 2012). Coming back to the results of the present study, the modulation of affective word processing by EUB suggests that differences between believers and non-believers in relation to emotional language are not due, at least exclusively, to consciously controlled processes such as the ones involved in a valence rating task. Importantly, these effects obtained with an on-line measure (i.e., RTs) may indicate the existence of individual differences by EUB in the organisation and dynamics of the networks involved in affective word processing. In this line, previous studies have suggested that individuals with unusual beliefs may present an increased emotional reactivity/sensitivity (e.g., Karcher & Shean, 2012; Kerns, 2005; Kerns & Berenbaum, 2000; van't Wout et al., 2004). This mechanism would fit with the results obtained in this study regarding valence since the effects of positive valence (facilitation) and negative valence (inhibition) became higher with increasing degree of EUB endorsement. However, it does not fit with the results obtained with arousal: under this hypothetical mechanism, we would also expect the effects of arousal to become higher with increasing levels of EUB, but what we obtained is precisely the inverse pattern. Therefore, the proposal of a heightened emotional reactivity/sensitivity is either a valence-specific mechanism, or it is not explaining the interactive effects observed here at all. Future studies could be oriented in trying to disentangle the underlying mechanisms behind the modulation of affective word processing by individuals' levels of EUB. Regardless of the explanatory mechanism, the interactive effects found in the present study are relevant in the context of the conflicting results of words' emotional content on visual word processing (see Hinojosa et al., 2020). Indeed, following a similar rationale as in the study of Silva et al. (2012) commented in the introduction, given the influence of EUB in affective word processing, differences in

https://doi.org/10.1017/langcog.2022.38 Published online by Cambridge University Press

the proportion/distribution of this variable across study samples may contribute, at least partially, to these inconsistencies. More specifically regarding valence, a study sample with either more believers or more non-believers in pseudoscience would foster or hinder, respectively, the appearance of a valence linear facilitating effect. With respect to arousal, a study sample with either more believers or more non-believers in superstition, occultism and/or pseudoscience would hinder or foster, respectively, the appearance of an arousal linear facilitating effect.

Regarding the effect of EUB on signal detection theory response pattern, we have replicated the association of high EUB with a more liberal response bias found in previous studies (e.g., Harrison et al., 2021; Krummenacher et al., 2010; Riekki et al., 2013; Rodríguez-Ferreiro & Barberia, 2021a). However, this bias was only observed in the recognition task, but not in the LDT. This task-dependence effect may arise from differences in the difficulty of each task (i.e., LDT is easier than the recognition task). In that sense, this liberal response bias may have been only activated in the recognition task, given the ambiguity/uncertainty derived from not being sure if the presented word was old or new. In contrast, it would not have been activated in the LDT because of its easiness. This would be congruent with all the evidence that links EUB with uncertainty and lack of control (see Douglas et al., 2020; French & Stone, 2014), and also with models that attribute to negative emotions, an activating/exacerbating role regarding EUB-related cognitive biases (e.g., Irwin, 2009; van Prooijen, 2020).

Paranormal, pseudoscientific and conspiracy beliefs have been grouped into the EUB category (Lobato et al., 2014; Rizeq et al., 2020), but it is not clear to what extent different instances of EUB share similar mechanisms: there have been both results that generalise across different instances of EUB [e.g., both paranormal and pseudoscientific believers seem to require a lower amount of evidence to draw conclusions than non-believers (Rodríguez-Ferreiro & Barberia, 2021b)] and others that do not [e.g., the degree of conspiracy belief predicted local-to-global and global-to-local interference effects in a visual attention paradigm, whereas the degree of paranormal belief was not a significant predictor (van Elk, 2015; see also Williams et al., 2022 for the suggestion that some cognitive biases associated with paranormal beliefs may be topic/domain specific]. In this context, even though the examined EUB instances in the present research (i.e., superstitious, occultist, religious, extraordinary creatures-related, conspiracy and pseudoscientific beliefs) share the common feature of being socially widespread beliefs in spite of not being epistemically grounded enough, our results suggest that they would have differential specificities in relation to the mechanisms underlying the effects studied here. Of note, caution should be taken in relation to the results with PEUBI-TR and PEUBI-ELF since the lack of variability (i.e., most of the participants scored low, as indicated by the high positive skew) may be problematic for inferential purposes (see Hallgren, 2018). Future studies should further explore the pattern of effects with more heterogeneous samples in relation to religious and extraordinary creatures-related beliefs.

In sum, the present study provides evidence about the role of subjects' individual differences in EUB on the processing of words' emotional content. It also adds to the literature that has found a liberal response criterion in EUB believers. Both effects seem not to be general to all EUB, which favours the idea that different instances of EUB have their specificities. These findings have several implications. First, from a psycholinguistic perspective, our results show that subjects' individual differences matter and, therefore, that they should be methodologically and theoretically considered in psycholinguistics. Second, regarding the basic psychological processes underlying EUB, this study provides evidence of the existence of individual differences by EUB in basic psycholinguistic processes such as affective word processing.

Acknowledgements. We would like to thank Juan Haro for the help given to the first author by introducing him to the programming language R and to LMEM. We would also like to thank Harry Price for the linguistic revision of the manuscript.

Supplementary materials. To view supplementary material for this article, please visit http://doi.org/10.1017/langcog.2022.38.

Data availability statement. The data and R scripts that support the findings of this study are openly available in the Open Science Framework (OSF) repository at https://osf.io/pe7u2/.

Funding statement. This work was supported by the Spanish Government (project PID2019-107206GB-I00 funded by MCIN/AEI/10.13039/501100011033 + DHP's predoctoral contract FPU20/03345 from the call *Ayudas para la formación de profesorado universitario – Convocatoria 2020*). Open access Article Processing Charges (APC) were paid with funds from the Psychology Department of Universitat Rovira i Virgili.

Competing interests. The authors report there are no competing interests to declare.

References

Adelman, J. S. (2012). Methodological issues with words. In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 116–138). Psychology Press.

Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Duñabeitia, J. A. (2018). SPALEX: A Spanish lexical decision database from a massive online data collection. *Frontiers in Psychology*, 9, 2156. https://doi.org/10.3389/fpsyg.2018.02156

Alonso, M. A., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods*, 47(1), 268–274. https://doi.org/10.3758/s13428-014-0454-2

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Ball, M. J., Perkins, M. R., Müller, N., & Howard, S. (2008). *The handbook of clinical linguistics*. Blackwell Publishing. https://doi.org/10.1002/9781444301007

Barriga-Paulino, C. I., Guerreiro, M., Faísca, L., & Reis, A. (2022). Does emotional valence modulate word recognition? A behavioral study manipulating frequency and arousal. *Acta Psychologica*, 223, 103484. https://doi.org/10.1016/j.actpsy.2021.103484

Barton, K. (2022). *MuMIn: Multi-model inference (R package version 1.46.0)* [Computer software]. https://cran.r-project.org/web/packages/MuMIn.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *lme4: Linear mixed-effects models using 'eigen' and S4 (R package version 1.1-29)* [Computer software]. https://cran.r-project.org/web/packages/lme4.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263–292. https://doi.org/10.1037/bul0000089

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings.* Technical report C-1, The Center for Research in Psychophysiology. University of Florida. https://pdodds.w3.uvm.edu/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf

Brugger, P., & Graves, R. E. (1997). Testing vs. believing hypotheses: Magical ideation in the judgement of contingencies. *Cognitive Neuropsychiatry*, 2(4), 251–272. https://doi.org/10.1080/135468097396270

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 9. https://doi.org/10.5334/joc.10

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, 18(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, 68(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology*, 3, 348. https://doi.org/10.3389/fpsyg.2012.00348

Douglas, K. M., Cichocka, A., & Sutton, R. M. (2020). Motivations, emotions and belief in conspiracy theories. In M. Butter & P. Knight (Eds.), *Routledge handbook of conspiracy theories* (pp. 181–191). Routledge.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Dyer, K. D., & Hall, R. E. (2019). Effect of critical thinking education on epistemically unwarranted beliefs in college students. *Research in Higher Education*, 60(3), 293–314. https://doi.org/10.1007/s11162-018-9513-3

Fasce, A., Avendaño, D., & Adrián-Ventura, J. (2021). Revised and short versions of the Pseudoscientific Belief Scale. *Applied Cognitive Psychology*, 35(3), 828–832. https://doi.org/10.1002/acp.3811

Feng, W. (2016). An approach of power estimation for linear mixed models for clinical studies. *Science Journal of Applied Mathematics and Statistics*, 4(2), 59–63. https://doi.org/10.11648/j.sjams.20160402.17

Ferré, P., Haro, J., Huete-Pérez, D., & Fraga, I. (2021). Emotionality effects in ambiguous word recognition: The crucial role of the affective congruence between distinct meanings of ambiguous words. *Quarterly Journal of Experimental Psychology*, 74(7), 1234–1243. https://doi.org/10.1177/17470218219900

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124. https://doi.org/10.3758/BF03195503

Fox, E. (2020). Readers' individual differences in affect and cognition. In E. B. Moje, P. Afflerbach, P. Enciso, N. K. Lesaux, & N. K. Lesaux (Eds.), *Handbook of reading research* (Vol. V, pp. 180–196). Routledge.

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., … R-Core Team (2022). *car: Companion to applied regression (R package version 3.0-13)* [Computer software]. https://cran.r-project.org/web/packages/car

Fox, J., Weisberg, S., Price, B., Friendly, M., Hong, J., Andersen, R., Firth, D., Taylor, S., & R-Core Team (2022). *effects: Effect displays for linear, generalized linear, and other models (R package version 4.2-1)* [Computer software]. https://cran.r-project.org/web/packages/effects

Fraga, I., Guasch, M., Haro, J., Padrón, I., & Ferré, P. (2018). EmoFinder: The meeting point for Spanish emotional words. *Behavior Research Methods*, 50(84), 1–10. https://doi.org/10.3758/s13428-017-1006-3

French, C. C., & Stone, A. (2014). *Anomalistic psychology: Exploring paranormal belief and experience*. Palgrave Macmillan.

Galasová, M. (2022). *It is easier with negative emotions: The role of negative emotions and emotional intelligence in epistemically suspect beliefs about COVID-19 [Conference paper].* Cognition and Artificial Life. https://www.researchgate.net/publication/360969568_It_Is_Easier_with_Negative_Emotions_The_Role_of_Negative_Emotions_and_Emotional_Intelligence_in_Epistemically_Suspect_Beliefs_about_COVID-19

Gianotti, L. R. (2003). *Brain electric fields, belief in the paranormal, and reading of emotion words* [Doctoral dissertation, University of Zurich]. https://doi.org/10.5167/uzh-163143

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, 45(3), 765–771. https://doi.org/10.3758/s13428-012-0296-8

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y

Hallgren, K. A. (2018). Restriction of range. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1431–1433). SAGE Publishing. https://doi.org/10.4135/9781506326139.n595

Harrison, A. W., Shou, Y., & Christensen, B. K. (2021). A cognitive model of delusion propensity through dysregulated correlation detection. *Schizophrenia Research*, 237, 93–100. https://doi.org/10.1016/j.schres.2021.08.025

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods*, 48(1), 272–284. https://doi.org/10.3758/s13428-015-0572-5

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neurolinguistics: Towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, 35(7), 813–839. https://doi.org/10.1080/23273798.2019.1620957

Hinojosa, J. A., Rincón-Pérez, I., Romero-Ferreiro, M. V., Martínez-García, N., Villalba-García, C., Montoro, P. R., & Pozo, M. A. (2016). The Madrid affective database for Spanish (MADS): Ratings of dominance, familiarity, subjective age of acquisition and sensory experience. *PLoS One*, 11(5), e0155866. https://doi.org/10.1371/journal.pone.0155866

Huete-Pérez, D., Morales-Vives, F., Gavilán, J. M., Boada, R., & Haro, J. (2022). Popular Epistemically Unwarranted Beliefs Inventory (PEUBI): A psychometric instrument for assessing paranormal, pseudo-scientific and conspiracy beliefs. *Applied Cognitive Psychology*, 36, 1260–1276. https://doi.org/10.1002/acp.4010

Irwin, H. J. (2009). *The psychology of paranormal belief: A researcher's handbook*. University of Hertfordshire Press.

Karcher, N., & Shean, G. (2012). Magical ideation, schizotypy and the impact of emotions. *Psychiatry Research*, 197(1–2), 36–40. https://doi.org/10.1016/j.psychres.2011.12.033

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing*, 25(6), 1259–1282. https://doi.org/10.1007/s11145-011-9316-9

Kerns, J. G. (2005). Positive schizotypy and emotion processing. *Journal of Abnormal Psychology*, 114(3), 392–401. https://doi.org/10.1037/0021-843X.114.3.392

Kerns, J. G., & Berenbaum, H. (2000). Aberrant semantic and affective processing in people at risk for psychosis. *Journal of Abnormal Psychology*, 109(4), 728–732. https://doi.org/10.1037/0021-843X.109.4.728

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, 22(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14–34. https://doi.org/10.1037/a0021446

Krummenacher, P., Mohr, C., Haker, H., & Brugger, P. (2010). Dopamine, paranormal belief, and the detection of meaningful stimuli. *Journal of Cognitive Neuroscience*, 22(8), 1670–1681. https://doi.org/10.1162/jocn.2009.21313

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143(3), 1065–1081. https://doi.org/10.1037/a0035669

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in linear mixed effects models (R package version 3.1–3)* [Computer software]. https://cran.r-project.org/web/packages/lmerTest

Lobato, E., Mendoza, J., Sims, V., & Chin, M. (2014). Examining the relationship between conspiracy theories, paranormal beliefs, and pseudoscience acceptance among a university population. *Applied Cognitive Psychology*, 28(5), 617–625. https://doi.org/10.1002/acp.3042

Lüdecke, D., Bartel, A., Schwemmer, C., Powell, C., Djalovski, A., & Titz, J. (2021). *sjPlot: Data visualization for statistics in social science (R package version 2.8.10)* [Computer software]. https://cran.r-project.org/web/packages/sjPlot

Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. https://doi.org/10.1016/j.jml.2020.104092

Mueller, C. J., & Kuchinke, L. (2016). Individual differences in emotion word processing: A diffusion model analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 16(3), 489–501. https://doi.org/10.3758/s13415-016-0408-5

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition. Volume 2: Meaning and context, individuals and development* (pp. 24–43). Psychology Press.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. https://doi.org/10.1017/s0954579405050340

Real Academia Española. (2014). *23a edición del Diccionario de la lengua española.*

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On pleasure and thrill: The interplay between arousal and valence during visual word recognition. *Brain and Language*, 134, 34–43. https://doi.org/10.1016/j.bandl.2014.03.009

Redondo, J., Fraga, I., Comesaña, M., & Perea, M. (2005). Estudio normativo del valor afectivo de 478 palabras Españolas. *Psicológica*, 26, 317–326. http://www.redalyc.org/articulo.oa?id=16926207

Riekki, T., Lindeman, M., Aleneff, M., Halme, A., & Nuortimo, A. (2013). Paranormal and religious believers are more prone to illusory face perception than skeptics and non-believers. *Applied Cognitive Psychology*, 27(2), 150–155. https://doi.org/10.1002/acp.2874

Rizeq, J., Flora, D. B., & Toplak, M. E. (2020). An examination of the underlying dimensional structure of three domains of contaminated mindware: Paranormal beliefs, conspiracy beliefs, and anti-science attitudes. *Thinking & Reasoning*, 27(2), 187–211. https://doi.org/10.1080/13546783.2020.1759688

Rodríguez-Ferreiro, J., & Barberia, I. (2021a, June 25). *Endorsement of unwarranted beliefs is associated with liberal response criterion in a false memory* [Poster presentation]. XV International Symposium of Psycholinguistics. https://actos.nebrija.es/58817/detail/xv-international-symposium-of-psycholinguistics-2021.html

Rodríguez-Ferreiro, J., & Barberia, I. (2021b). Believers in pseudoscience present lower evidential criteria. *Scientific Reports*, 11(1), 1–7. https://doi.org/10.1038/s41598-021-03816-5

Rodríguez-Ferreiro, J., & Davies, R. (2019). The graded effect of valence on word recognition in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(5), 851–868. https://doi.org/10.1037/xlm0000616

Rossi, E., & Diaz, M. T. (2016). How aging and bilingualism influence language processing: Theoretical and neural models. *Linguistic Approaches to Bilingualism*, 6(1–2), 9–42. http://doi.org/10.1075/lab.14029.ros

Rotello, C. M. (2017). Signal detection theories of recognition memory. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 201–225). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21044-4

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12(2), 347–367. https://doi.org/10.1177/1094428107308

Silva, C., Montant, M., Ponz, A., & Ziegler, J. C. (2012). Emotions in reading: Disgust, empathy and the contextual learning hypothesis. *Cognition*, 125(2), 333–338. https://doi.org/10.1016/j.cognition.2012.07.013

Singmann, H., & Kellen, D. (2020). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Psychology Press. https://discovery.ucl.ac.uk/id/eprint/10107874/1/singmann_kellen-introduction-mixed-models%281%29.pdf

Stadthagen-González, H., Imbault, C., Pérez-Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, 49(1), 111–123. https://doi.org/10.3758/s13428-015-0700-2

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. https://doi.org/10.3758/BF03207704

Tárrega, J., Perea, M., Rojo-Bofill, L. M., Moreno-Giménez, A., Almansa-Tomás, B., Vento, M., & García-Blanco, A. (2021). Do children with overweight respond faster to food-related words? *Appetite*, 161, 105134. https://doi.org/10.1016/j.appet.2021.105134

Tremblay, A., & Ransijn, J. (2020). *LMERConvenienceFunctions: Model selection and post-hoc analysis for (g) lmer models (R package version 3.0)* [Computer software]. https://cran.r-project.org/web/packages/LMERConvenienceFunctions

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2), 302–324. https://doi.org/10.1075/ml.6.2.04tre

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, 39(4), 973–978. https://doi.org/10.3758/BF03192992 van Elk, M. (2015). Perceptual biases in relation to paranormal and conspiracy beliefs. *PLoS One*, 10(6), e0130422. https://doi.org/10.1371/journal.pone.0130422

van Prooijen, J. W. (2020). An existential threat model of conspiracy theories. *European Psychologist*, 25(1), 16–25. https://doi.org/10.1027/1016-9040/a000381

van–t Wout, M., Aleman, A., Kessels, R. P., Larøi, F., & Kahn, R. S. (2004). Emotional processing in a non-clinical psychosis-prone sample. *Schizophrenia Research*, 68(2–3), 271–281. https://doi.org/10.1016/j.schres.2003.09.006

Veldhuis, D., & Kurvers, J. (2012). Offline segmentation and online language processing units: The influence of literacy. *Written Language & Literacy*, 15(2), 165–184. https://doi.org/10.1075/wll.15.2.03vel

Wickham, H., Bryan, J., RStudio, Kalicinski, M., Valery, K., Leitienne, C., Colbert, B., Hoerl, D., & Miller, E. (2022). *readxl: Read excel files (R package version 1.4.0)* [Computer software]. https://cran.r-project.org/web/packages/readxl

Williams, C., Denovan, A., Drinkwater, K., & Dagnall, N. (2022). Thinking style and paranormal belief: The role of cognitive biases. *Imagination, Cognition and Personality: Consciousness in Theory, Research, and Clinical Practice*, 41(3), 274–298. https://doi.org/10.1177/02762366211036435

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/9781315165547

Yap, M., & Balota, D. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 26–43). Oxford University Press.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 53–79. https://doi.org/10.1037/a0024177

Yu, A. C., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, 5, 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

Cite this article: Huete-Pérez, D. & Ferré, P. (2023). Individual differences in visual word recognition: the role of epistemically unwarranted beliefs on affective processing and signal detection *Language and Cognition* 15: 314–336. https://doi.org/10.1017/langcog.2022.38