



# As easy as cake or a piece of pie? Processing idiom variation and the contribution of individual cognitive differences

Gareth Carrol<sup>1</sup> · Katrien Segaert<sup>2</sup>

Accepted: 30 August 2023 / Published online: 19 September 2023  
© The Author(s) 2023

## Abstract

Language users routinely use canonical, familiar idioms in everyday communication without difficulty. However, creativity in idiom use is more widespread than sometimes assumed, and little is known about how we process creative uses of idioms, and how individual differences in cognitive skills contribute to this. We used eye-tracking while reading and cross-modal priming to investigate the processing of idioms (e.g., *play with fire*) compared with creative variants (*play with acid*) and literal controls (*play with toys*), amongst a group of 47 university-level native speakers of English. We also conducted a series of tests to measure cognitive abilities (working memory capacity, inhibitory control, and processing speed). Eye-tracking results showed that in early reading behaviour, variants were read no differently to literal phrases or idioms but showed significantly longer overall reading times, with more rereading required compared with other conditions. Idiom variables (familiarity, decomposability, literal plausibility) and individual cognitive variables had limited effects throughout, although more decomposable phrases of all kinds required less overall reading time. Cross-modal priming—which has often shown a robust idiom advantage in past studies—demonstrated no difference between conditions, but decomposability again led to faster processing. Overall, results suggest that variants were treated more like literal phrases than novel metaphors, with subsequent effort required to make sense of these in the way that was consistent with the context provided.

**Keywords** Idioms · Metaphor · Individual differences · Eye-tracking · Cross-modal priming

Idioms, such as *a piece of cake* (meaning “really easy”), play an important role in theories of how natural language is used, processed, and learned. As (variably) noncompositional multiword expressions, they show how meaning can simultaneously be distributed over words and longer phrases; as figurative expressions, they help us understand how nonliteral meaning is dealt with; and as inherently ambiguous phrases, they show how the language processor uses multiple constraints to settle on an appropriate meaning. The current study aims to extend our understanding of how idioms are processed by focusing on two aspects that have received less attention in the existing literature: the ways in which idiomatic creativity affects processing and the

contribution of individual variability in cognitive abilities to how idioms are understood.

Existing research has led to the following broad findings:

- i) On the whole, idioms are recognized and understood more quickly than matched literal phrases by native speakers of a language (Gibbs, 1980; McGlone et al., 1994; Rommers et al., 2013; Swinney & Cutler, 1979). This is referred to throughout the literature as the “idiom superiority effect.”
- ii) Familiarity is a key driver of idiom recognition and understanding (Carrol et al., 2018; Columbus et al., 2015; Libben & Titone, 2008; Schweigert, 1986, 1991; Schweigert & Moates, 1988; Titone & Libben, 2014). In turn, familiar idioms represent highly predictable completions, which also contributes to the fast activation of figurative meaning (Cacciari & Tabossi, 1988; Tabossi et al., 2005; Tabossi & Zardon, 1993, 1995; Titone & Connine, 1994a).
- iii) Whilst initial recognition seems to be largely driven by familiarity, other factors are known to affect the selec-

✉ Gareth Carrol  
g.carrol@bham.ac.uk

<sup>1</sup> Department of English Language and Linguistics, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK

<sup>2</sup> School of Psychology and Centre for Human Brain Health, University of Birmingham, Birmingham, UK

tion and integration of meaning (figurative vs. literal), principally decomposability, or how much the individual words contribute to the meaning of the whole phrase (Caillies & Butcher, 2007; Titone & Connine, 1999; Titone & Libben, 2014), and literal plausibility, or whether the phrase could reasonably be interpreted literally (Cronk & Schweigert, 1992; Findlay & Carrol, 2019; Mueller & Gibbs, 1987; Titone & Libben, 2014). Individual differences are also assumed to affect this, since these (and other) factors will vary substantially across language users.

Several important caveats must be added. One is that the advantage of familiarity seems to extend to a range of “formulaic” phrases, including binomials, collocations, lexical bundles, and other frequently occurring multiword units (Arnon & Snider, 2010; Bannard & Matthews, 2008; Carrol & Conklin, 2020; Siyanova-Chanturia et al., 2011; Sonbul, 2015; Tabossi, Fanari, et al., 2009; Tremblay et al., 2011; Vilkaite, 2016). Second, processing is not “all or nothing,” and many researchers have adopted a “hybrid” position whereby idioms are subject to direct retrieval and compositional analysis simultaneously (Sprenger et al., 2006; Titone & Connine, 1999). Titone et al. (2019) demonstrated that the earliest stages of idiom processing seem to be driven chiefly by direct retrieval, whereas later stages of meaning integration demonstrate a mix of retrieval and compositional analysis, where factors such as decomposability and context interact to determine how the idiom will be understood. Lastly, despite a widespread view of idioms as “prototypically” formulaic (Siyanova-Chanturia & Martinez, 2014), corpus-based work has demonstrated that flexibility and variability are common (e.g., Fellbaum, 2019; Feyaerts, 2006; Langlotz, 2006a, 2006b; Moon, 1998; Vrbinc & Vrbinc, 2011; Wulff, 2008). How idiom variability is processed is a key question that we aim to address in the present study, with the expectation that modification (e.g., through changing component words) should require longer processing compared with canonical forms, but that individual differences in cognitive skills may affect how easily participants can retrieve the intended figurative meaning. We also aim to compare two methods that have been fruitfully applied in the literature, although rarely in tandem: eye tracking while reading and cross-modal priming.

## Idiom creativity

As mentioned above, corpus-based work has demonstrated that idiom flexibility and variability are common. Modifications to idioms often demonstrate conscious wordplay (Duffley, 2013) or explicit awareness of the citation form (Fellbaum, 2019), and a consensus is that idioms remain

retrievable, provided they at least partly maintain the underlying metaphorical sense of the original (Feyaerts, 2006; Langlotz, 2006a, 2006b). These studies show that idioms can be used in a range of creative ways but tell us little about how these are processed by the intended audience. Investigations into the processing of idiomatic variants have often focused on the acceptability of syntactically and lexically modified forms. For example, Tabossi, Wolf, et al. (2009) used acceptability judgments to show that idiom syntax is determined by similar principles as literal language, with supporting context increasing the acceptability of modified forms (for similar arguments about the role of context in supporting variability, see also Fellbaum, 2019; Glucksberg, 2001; Hovhannisyanyan & Mkrtchyan, 2014).

Other studies have confirmed that syntactic variability is generally accepted, although the extent of this varies considerably across idioms (Tabossi et al., 2011). In online processing, some transformations (e.g., passivisation: *spill the beans* → *the beans were spilled*) do not prevent idioms from being recognized (Kyriacou et al., 2020) or the figurative meaning being activated (Mancuso et al., 2020), whereas others (e.g., adjective insertion: *spill the spicy beans*) lead to significantly longer reading times, indicating more effortful processing (Haeuser et al., 2020; Kyriacou et al., 2021). Of note here, when studies have looked at processing across different stages (e.g., comparing early vs. late eye-tracking measures), effects are much more consistently observed in later measures than during earlier stages of processing (Haeuser et al., 2020; Kyriacou et al., 2020, 2021). This in turn suggests that while initial reading of noncanonical idioms may be relatively straightforward (possibly because they were treated as literal phrases), later stages, where compositional analysis is likely to demonstrate that a literal reading is anomalous, are where slowdown is more likely to occur.

Lexical variation is generally more disruptive than syntactic variation, but the extent of this can be affected by familiarity with the underlying idiom (McGlone et al., 1994), context (Holsinger, 2013), transparency (Geeraert et al., 2017), and how close the replacement word is to the original idiom (Hamblin & Gibbs, 1999), as shown through a range of rating, reading, and visual-world eye-tracking tasks.<sup>1</sup> Smolka and Eulitz (2020) used acceptability ratings to compare idioms (*reach for the stars*) and variants (*reach for the planets*). Ratings for variants were lower than for canonical idioms but higher than for unrelated control

<sup>1</sup> In these studies, lexical variation took different forms, including replacement with a near synonym (*kick the bucket* → *kick the pail*; *kick the bucket* → *punt the bucket*), insertion of an additional element (*spill the beans* → *spill a single bean*), and blends of similar idioms (*go through the roof* + *off the charts* → *go through the charts*).

forms. They also found that modification of nouns led to lower acceptability ratings than modification of either verbs or prepositions.

A related line of research has shown that even when translated from another language, idioms remain recognizable and interpretable (Carrol et al., 2016; Carrol & Conklin, 2014, 2017; Senaldi et al., 2022; Titone et al., 2015), suggesting that a specific surface form is not the only possible route to idiom processing. Conversely, when idioms are entirely unfamiliar, processing is shown to be highly disrupted as language users must actively engage in the process of working out the intended meaning. Carrol and Littlemore (2020) showed that entirely unknown idioms came with a predictable cost during natural reading and were read significantly more slowly than literal paraphrases. Importantly, more transparent phrases had shorter reading times in a way that was not seen for known idioms or conventional metaphors (see also Libben & Titone, 2008, for evidence that decomposability has a limited effect on highly familiar idioms). The literature on metaphor processing also shows that factors such as familiarity and conventionality are primary drivers of understanding (Blasko & Briehl, 1997; Inhoff et al., 1984), while aspects relating to interpretability only play a role in less familiar phrases (Blasko & Connine, 1993).<sup>2</sup>

### Individual cognitive components of idiom processing

As in language processing more generally (e.g., Kidd et al., 2018), interest in the role of individual cognitive differences has provided a further set of questions to develop our understanding of how figurative language is understood. Evidence here has come from two sources: neuropsychological studies (of e.g., language impaired individuals) and studies measuring the impact of individual components of cognitive function on various language tasks.

For idioms, a series of investigations by Papagno and colleagues (Oliveri et al., 2004; Papagno et al., 2003, 2004, 2006; Papagno & Caporali, 2007; Papagno & Genoni, 2004) highlighted the importance of executive function (the suite of abilities that underpins the brain's ability to monitor and manage cognitive processes in a flexible, adaptable way) in processing figurative phrases amongst patients with Alzheimer's disease and aphasia. Executive function may play a vital role here since successful idiom interpretation requires

the language user to consider and disambiguate competing meanings, maintain relevant information in working memory, inhibit irrelevant meanings, and so on. In general, impairment led to more literal interpretations of idioms, and the authors suggested that this was due in part to an impaired ability to ignore the (less relevant) literal meanings in the contexts provided. For example, patients with mild Alzheimer's disease performed poorly on idiom comprehension during a sentence-to-picture matching task when a figurative versus literal interpretation of an idiom was offered, but performance improved significantly when the same task offered a figurative versus unrelated picture choice, and when asked to provide verbal explanations of idiom meanings (Papagno et al., 2003)

Amongst nonclinical participants, studies have largely focused on the role of working memory, and how this affects the ability to maintain and select between competing figurative and literal meanings. In metaphor studies, high working memory has been consistently linked to faster and more successful interpretation (Blasko, 1999; Carriedo et al., 2016; Chiappe & Chiappe, 2007; Kazmerski et al., 2003; Olkonemi et al., 2016; Pierce et al., 2010). Columbus et al. (2015) also looked at the role of executive control, focusing on individual differences amongst healthy participants and their effect on processing of metaphors and idioms. Their results suggest a pattern whereby higher executive control led to longer initial reading times for metaphors, indicating a tendency for readers to take information into account and commit to a particular interpretation of a potentially metaphorical verb as they were reading it. In comparison, participants with lower levels of executive control spent less time overall on metaphors (compared with literal controls) but were more likely to regress to a preceding context to help with interpretation. Idiom processing showed no such effects of executive control, supporting the view that they were retrieved holistically (with no cognitive control required) rather than analyzed/constructed online.

Cacciari et al. (2018), in an exploratory study, found that online idiom comprehension (response times in a cross-modal lexical decision task) was positively affected by higher scores for working memory, inhibition control and crystallized verbal intelligence. They argue that successful processing of idioms requires activation of multiple sources of information (literal meanings of words as well as the overall idiomatic meaning) as well as an ability to suppress the literal meaning once this is shown to be irrelevant. The authors also suggest that, as with language acquisition and processing more generally, a large vocabulary is required for this to be completed efficiently. To summarize the findings so far, links between working memory and metaphor processing in healthy adult language users have been demonstrated in several previous studies, with effects in both early and later measures of processing, while evidence for links

<sup>2</sup> Blasko and Connine (1993) include "aptness" in their analysis, which is functionally equivalent to what others have called transparency or motivation, in that it asks how well a particular metaphor expresses its intended figurative meaning.

between working memory or executive/inhibitory control and idiom processing are still tentative. An open question is how modified idioms of the type investigated in the studies discussed earlier in this section might therefore be affected by these variables, since they are neither entirely novel (like metaphors) nor entirely familiar (like idioms).

In the current study we focus on idiom variation, and the contribution of individual differences in cognitive function to how this is processed. Only one previous study has investigated this issue, albeit indirectly. Haeuser et al. (2020), as discussed in the previous section, compared processing of canonical versus modified (inserted adjective) idioms. They also compared idiom processing for younger and older adults, to compare contributions of fluid cognitive functions (working memory, processing speed, inhibition control) with crystallized cognitive abilities such as better world and lexical knowledge (including a more entrenched knowledge of idioms).<sup>3</sup> They found evidence that older adults accessed canonical idioms more quickly than younger participants (relative to literal controls), and that subsequent literal biasing context caused more difficulty for older readers, which they interpreted as indicating difficulty in suppressing a figurative meaning when this was necessary. Modified forms of idioms caused similar problems for older and younger participants, with both groups showing a bias toward a literal reading. As above, effects of modification were most apparent in later measures of processing, which is where the various idiom and cognitive factors discussed so far may have the biggest role to play (as opposed to earlier, “retrieval-based” stages—Titone et al., 2019).

## Summary and hypotheses

The literature discussed above broadly reveals a picture where creative variation does not necessarily block access to idiom meanings but does lead to longer processing times, assumed to represent the more effortful process of reconfiguring the phrase to reach the intended meaning. The aim of the current study is to investigate the online processing of idioms and creative modifications. We predict that modifying idioms by changing the final noun will lead to longer processing times, but that individual differences in cognitive variables such as working memory and inhibitory control may affect how easily modifications are processed. Similarly, individual properties of idioms may attenuate this process. More decomposable idioms may be more amenable to

modification, since the underlying figurative meaning should be less disrupted than for less decomposable phrases. If variants are initially treated as literal phrases, then these effects (slower processing for modified forms but also the influence of both cognitive and idiom variables) should be most apparent during later stages of processing, where compositional analysis and meaning integration are assumed to play a more prominent role. Conversely, if variants are immediately recognized as either modifications of known phrases or as novel metaphors in their own right, the earliest lexical recognition stages may also be impacted. Individual familiarity may also reduce disruption, if highly familiar idioms can be more easily accessed and reconfigured, or, conversely, may serve to make modification more disruptive, if more familiar idioms are more firmly entrenched. Effects here may show up in both early (idiom recognition) and late (reconfiguration and analysis) stages of processing.

## Method

The study uses two of the principal behavioural methods for investigating idioms: eye tracking while reading and cross-modal priming (see Carrol, 2021, for a review of these methods in figurative language research). Eye-tracking records eye movements during natural reading, with both duration and sequence of fixations contributing to an understanding of how text is processed. Fixation measures are often divided into “early,” reflecting the initial processes of word recognition/contact with the lexicon, and “late,” reflecting the process of meaning selection/integration (Altarriba et al., 1996; Conklin et al., 2018; Godfroid, 2020; Inhoff, 1984; Paterson et al., 1999; Rayner, 2009). This provides a way to infer the ease or difficulty with which idioms are processed and has helped to show the difference between recognizing a known phrase (through early measures such as first pass reading time) and making sense of the meaning (through later measures such as total reading time and regressions back into an idiom).

Cross-modal priming (Swinney, 1979; Swinney et al., 1979) has been used to demonstrate the activation of figurative or literal meaning during the processing of idioms. Participants listen to a spoken sentence and then at a key point are presented with a visual target word and asked to make a lexical decision (Is this a real word, yes or no?) as quickly as possible. In idiom studies, targets are typically related to figurative or literal meanings of the idiom, and compared with a baseline (an unrelated word) to assess the degree and/or time course of activation. A common finding is that responses to figuratively related words are faster than to unrelated words, which is taken as evidence for the rapid activation of idiom meanings for well-known phrases. These two methods have each been fruitfully applied in the study of

<sup>3</sup> Although Haeuser et al. (2020) report the results of a range of cognitive tests to demonstrate differences between their participant groups, they did not directly use these scores as predictors in their main analysis.

**Table 1** Characteristics of stimuli, including frequency (on the Zipf scale of 1–7<sup>1</sup>) of final nouns for idioms, variants and literal phrases; decomposability ratings (1–5) for idioms (Idiom Decomp); semantic

similarity scores (0–1) between idioms and variants; and substitutability ratings (1–5) for idioms and variants

	Idiom Noun freq	Variant Noun freq	Literal Noun freq	Idiom Decomp	Semantic Similarity	Substitutability
Mean ( <i>SD</i> )	4.67 (0.57)	4.65 (0.53)	4.67 (0.52)	3.06 (0.77)	0.54 (0.27)	3.04 (0.69)
Range	2.83–5.63	2.57–5.51	3.20–5.58	1.64–4.62	0.13–1.00	1.4–4.5

<sup>1</sup>The Zipf scale (van Heuven et al., 2014) is a logarithmic scale reflecting relative frequency given the size of the corpus being used. A value of 1 represents 1 occurrence per 100 million words, 2 represents 10 occurrences per 100 million words, 3 represents 100 occurrences, and so on.

idioms but are rarely used together, and offer different perspectives on this issue. Whilst cross-modal priming offers a way to directly measure activation of meaning, eye tracking provides a more inferential account whereby we can assess the factors that make processing in context easier or harder. Our experimental session therefore incorporated a reading study and a cross-modal lexical decision task, to directly compare results from the two.

## Participants

Forty-seven participants took part (42 females, mean age = 19.3 years, *SD* = 1.3) and everyone completed all aspects of the study. Participants were all monolingual native speakers of English, with normal or corrected-to-normal vision and no self-reported history of speech or language disorders. Half of the participants completed the reading task first and half completed the cross-modal priming task first. In between the reading and cross-modal priming task, a series of cognitive tests was administered, and following all tasks a familiarity test was administered to assess how well each participant knew the idioms they had seen during the study. Overall, participants took around 1 hour to complete all tasks.

## Materials

Fifty-one idioms of the form “verb-X-noun” (e.g., *play with fire*) were selected from previously published studies (see the Supplementary Materials for a detailed description of item selection). We chose to investigate lexical variation as this has had the clearest impact on idiom comprehension, based on previous work. For all items we created a plausible variant by replacing the final noun in a way that maintained the metaphorical sense of the idiom. For example, in *play with fire*, we replaced *fire* with *acid*, to maintain the underlying meaning of *play with [dangerous thing]*. A literal control phrase for each idiom was also created in the same way, hence for *play with fire* the control was *play with toys*. Nouns were matched as closely as possible for length and individual frequency (from the British National Corpus). Ratings for literal plausibility, decomposability, semantic similarity

between idiom and variant nouns, and acceptability of variants were collected for all items. Table 1 summarizes the characteristics of the stimuli, and a full list of experimental items is provided in the Appendix.

For the reading study, we created context sentences that supported either the idiom or variant meaning (same context for both) or the literal phrase. All sentences consisted of two conjoined clauses. The first clause always began with a personal pronoun and introduced the target phrase in a neutral way, followed by the target phrase (e.g., *They have both been playing with [fire / acid / toys]*). The second clause followed a conjunction and was consistent with the figurative (e.g., *and it's no surprise that things have gone badly*) or literal (e.g., *and seem to have been getting on very well together*) version of the phrase. Fifty-one filler items were created, comprising literal sentences of approximately the same length and syntactic structure as the critical sentences. Items were counterbalanced over three presentation lists, with an idiom, its variant, and its literal control phrase appearing on different lists. Each list contained 17 items in each condition, and the same filler sentences were used in all three presentation lists, for a total of 102 items per list.

For the cross-modal priming, a word related to the figurative meaning of the idiom was chosen (e.g., for *play with fire* the target was “danger”), matched across all items for length and frequency. The sentences created for the reading task were all reused for the cross-modal priming task, up to the end of the key phrase (e.g., for *play with fire* the stimulus item in the idiom condition was *They have both been playing with fire*). Filler items from the reading task were adapted by cutting the sentence off after the first clause, to leave items of around the same length as the critical stimuli. Filler items were always paired with a pseudoword. We also added in an additional 17 unused idioms, to give a total of 68 fillers. The purpose of this was to ensure that some (noncritical) idioms could be followed by a pseudoword to avoid participants developing expectancy-based strategies. The critical and filler items were used to create sound files using an online voice synthesizer recorded with a female voice with a British English accent. Critical items were paired with their related word and counterbalanced over the same three presentation lists as in the reading study, meaning that 17 idioms (critical

items), 17 variants, and 17 literal controls were followed by a real word, and 68 fillers (51 literal sentences and 17 idioms) were followed by a pseudoword. Pseudowords were all pseudohomophones between four and eight letters long, created using the ARC Nonword Database (Rastle et al., 2002).

Participants were randomly assigned to one of the combinations of presentation lists (AB, AC, BC) so that no one saw the same list for the two main tasks—that is, if a participant saw *play with fire* in the reading task, they would see either *play with acid* (variant) or *play with toys* (literal) in the cross-modal priming task. Whilst the use of two tasks with overlapping stimuli opens up the possibility of priming effects between the two, the use of counterbalanced presentations lists should ensure that these effects were minimized.

### Reading task

Participants were seated in front of a desk-mounted Eye-Link 1000+ eye tracker (SR Research). A chin rest was used to ensure head stability during the task. Following a nine-point calibration and validation procedure, participants were told that they would be presented with a series of single sentences that they should read as naturally as possible for comprehension. Recording was monocular at 500 Hz. Each trial was preceded by a fixation point, then the sentence was displayed in the middle of a white screen (47.5cm × 27cm, 1,680 × 1,050 pixels) in Courier New font size 17 pt. Forty items (10 per condition and 10 fillers) were followed by a simple yes/no comprehension question. Following six practice items, all 102 trial sentences were presented in random order. A self-timed break was offered midway through, and recalibrations were performed after the practice items, at the mid-point, and as required during the session.

### Cross-modal priming

The cross-modal priming study was built and run in E-Prime (Version 2.0). Participants were told that they would hear a spoken sentence played through speakers and immediately afterward would see a word appear on the screen. They should indicate as quickly as possible whether what they saw was a real English word using buttons marked Y and N on a serial response button box. The target word appeared 250 ms after the offset of the spoken phrase in the centre of a white screen (47.5cm × 27cm, 1,680 × 1,050 pixels) in Courier New font, size 18 pt. Based on previous studies, it was expected that idiom meaning was likely to be available immediately at the offset of the phrase for familiar items, but that figurative meaning for variants may take some time to emerge, hence a short delay was included. Participants saw eight practice items of the same form as the experimental stimuli, with a mix of word and pseudoword targets, then all items were presented in random order.

### Cognitive tests

We chose three aspects of cognitive function, based on the literature discussed previously: inhibitory control, processing speed, and working memory. Whilst working memory and inhibitory control have been included in previous idiom studies (as outlined earlier), we included processing speed in order to obtain a general idea of how quickly participants were able to process information, and how this varied in the different conditions. The cognitive tests were administered following the first task, in order to provide a distraction before participants began the second task (and to further reduce the risk of priming effects between the two main tasks). All participants completed the cognitive tasks in the same order. First, they were asked to complete a flanker task (which measures inhibitory control), programmed and delivered in E-Prime (Version 2.0). Here, participants are shown a series of five chevrons (>) pointing left or right and must indicate which way the middle one is pointing using buttons marked Left and Right on a serial response box. Trials are either congruent (where all point the same way, e.g., > > > > >) or incongruent (where the middle chevron points in a different direction, e.g., < < > < <). Performance on the incongruent trials is taken to reflect the ability to inhibit irrelevant information (the arrows pointing in the opposite direction), and the overall score is calculated as the difference in performance between congruent and incongruent trials, where a smaller difference indicates greater inhibition ability.

Second, a processing speed test was administered, using a version of the digit symbol substitution test (DSST) taken from the Wechsler Adult Intelligence Scale, Fourth UK Edition (WAIS-IV UK: Wechsler, 2010). This task presented participants with the numbers from 0 to 9 at the top of a page, along with a key assigning a symbol to each number. The page presents 135 numbered boxes and asks participants to draw the correct symbol for each, using the key provided. Following some demonstration items, participants have 2 minutes to complete as much of the test as possible, and the number of items correctly completed is taken as a measure of processing speed (max = 135).

Finally, participants were asked to complete two tests of working memory: a backward digit span task, where participants are required to repeat a sequence of numbers back in reverse order (e.g., if given “1–2”, a participant would answer “2–1”); and a subtract two task, where participants are given a sequence of numbers and must subtract two from each digit in the sequence (e.g., if given “5–3”, a participant would answer “3–1”). In both tasks, the size of the sequence increases until participants can no longer reliably provide the required numbers, and the overall working memory span is calculated as the maximum sequence size for which they could complete at least three examples out of a block of five.

**Table 2** Inhibition score (difference between congruent and incongruent flanker trials, in ms, with scores closer to 0 representing better inhibition), working memory (average of the two tests showing the longest sequence successfully repeated), processing speed (coding score out of 135), and familiarity score (average out of 5, including a score of 0 for unknown phrases) for all participants

	Inhibitory control (flanker score)	Working memory (average)	Processing speed (coding score)	Familiarity
Mean ( <i>SD</i> )	103 (42)	3.6 (0.8)	83 (12)	3.9 (0.6)
Range	35–248	2.3–6.0	56–110	2.4–4.9

For inhibitory control, all incorrect responses and RTs of 0 ms were removed; no response above 2,000 ms was recorded, and this was taken to be an acceptable upper limit.

Overall working memory was calculated as the average of the two scores for each participant.

### Familiarity test

Finally, participants completed a familiarity test to indicate how well they knew the items used in the main studies. All 51 idioms were presented in their citation forms and participants were asked to indicate on a 6-point scale whether they knew the idiom (where 0 = *I have never heard this idiom before*) and, if so, how familiar they were with it (where 1 = *I have heard it but I'm not sure what it means*, and 5 = *I know this idiom very well and am entirely sure of what it means*).

## Results

Eye-tracking data was unusable due to technical issues or consistently poor calibration for six participants, leaving a total of 41 subjects for the reading task and 47 subjects for the cross-modal priming task. Cognitive test and familiarity data was available for all subjects and is summarized in Table 2. No significant correlations existed amongst the four variables.

We analyzed the eye-tracking and cross-modal priming data separately, using linear mixed-effects models in R (Version 4.2.1) and RStudio (Version 2002.07.0) using the packages lme4 (Version 1.1-30) and lmerTest (Version 3.1-3). We applied the following analyses for each study:

1. Initial analysis to compare patterns between conditions. Condition was treatment coded with literal phrases as the baseline to compare deviation from this for idioms and variants. We then used the emmeans package (with Bonferroni adjustment for multiple comparisons) to compare idiom and variant conditions directly.

2. Inclusion of the idiom variables familiarity, decomposability, and literal plausibility.
3. Inclusion of the cognitive variables inhibitory control, processing speed, and working memory.

In all analyses, the data were collapsed across lists, and all continuous variables were centred and scaled prior to inclusion. We retained trial order and length and frequency (of final nouns for the reading data and of target words for the cross-modal priming data) as covariates in all models.

### Eye-tracking analysis

Overall accuracy on comprehension questions was 90%, indicating that participants had paid attention throughout. Eye-tracking data were checked visually for accuracy, and trials where track loss was evident or where the whole of the critical phrase received no fixations were removed. This led to the loss of 3.4% of the data, leaving 2,019 trials for analysis. Since all reading took place on a single line, data were also vertically aligned to correct any minor deviations in terms of the y-position. The data were then cleaned according to the four-stage process in the Eyelink Data-Viewer software (SR Research), which removes extreme values (fixations below 80 ms and over 800 ms). Any item where a participant had indicated that they did not know the underlying idiom (score of 0 on the familiarity task) was also removed. This led to the removal of a further 159 trials (7.9% of the usable data), leaving 1,860 trials for analysis.

Table 3 presents reading data for literal control phrases, idioms, and variants. The data summarises a range of duration (first-pass reading time, total reading time, regression path duration), skipping (likelihood of skipping the final noun), and regression measures (percentage of trials with a regression into or out of a given region) for three areas of interest: the whole of the critical phrase, the final noun of each phrase, and the post-phrase region (everything following the critical phrase, to the end of the sentence).

We constructed initial models for all of the variables listed in Table 3. Duration measures were log transformed prior to analysis, and for any models involving binary variables (final word skip and regressions in), a generalized linear model with binomial distribution was used. All standard models included the fixed effect of condition and the covariates described above, and included random intercepts for subject and item, and by-subject and by-item random slopes for the effect of condition.<sup>4</sup> We also checked substitutability scores (how acceptable variants were in conveying the

<sup>4</sup> Where this led to convergence issues, we removed random correlations between intercepts and slopes for items, then for subjects. All models converged after this.

**Table 3** Summary of reading data for literal control phrases, idioms, and variants

	Phrase		Final noun		Idiom		Variant		Literal		Post-phrase	
	Literal	Idiom	Literal	Variant	Literal	Variant	Idiom	Variant	Literal	Variant	Idiom	Variant
Skip rate %	-	-	0.22 (0.42) [0.19, 0.25]	-	0.22 (0.41) [0.19, 0.25]	0.18 (0.39) [0.15, 0.22]	-	-	-	-	-	-
First-pass RT (gaze duration)	458 (219) [442, 476]	431 (189) [416, 447]	190 (113) [181, 198]	463 (231) [446, 482]	183 (113) [174, 192]	210 (114) [201, 219]	1403 (646) [1310, 1460]	1364 (643) [1310, 1420]	1364 (643) [1310, 1420]	1362 (685) [1310, 1420]	1403 (646) [1350, 1460]	1362 (685) [1310, 1420]
Total RT	596 (290) [574, 619]	550 (252) [531, 571]	220 (150) [208, 232]	670 (333) [646, 697]	204 (140) [193, 215]	267 (192) [252, 283]	1564 (684) [1510, 1620]	1598 (690) [1550, 1660]	1598 (690) [1550, 1660]	1650 (670) [1600, 1710]	1564 (684) [1510, 1620]	1650 (670) [1600, 1710]
Regression path (go-past time)	595 (362) [568, 624]	566 (354) [541, 596]	266 (282) [246, 291]	625 (359) [600, 656]	236 (211) [221, 254]	291 (255) [272, 312]	1692 (684) [1630, 1760]	1758 (690) [1690, 1830]	1758 (690) [1690, 1830]	1863 (670) [1800, 1940]	1692 (684) [1630, 1760]	1863 (670) [1800, 1940]
Regressions out %	0.18 (0.38) [0.15, 0.21]	0.17 (0.38) [0.14, 0.20]	0.17 (0.37) [0.14, 0.20]	0.21 (0.41) [0.18, 0.25]	0.15 (0.36) [0.12, 0.18]	0.20 (0.40) [0.17, 0.23]	0.27 (0.44) [0.24, 0.30]	0.31 (0.46) [0.28, 0.35]	0.31 (0.46) [0.28, 0.35]	0.27 (0.44) [0.24, 0.30]	0.27 (0.44) [0.24, 0.30]	0.39 (0.49) [0.35, 0.43]
Regressions in %	0.18 (0.39) [0.15, 0.21]	0.14 (0.35) [0.11, 0.17]	-	0.23 (0.42) [0.20, 0.26]	-	-	-	-	-	-	-	-

Duration measures are in milliseconds and figures for final nouns include zeroes for items skipped during first-pass reading; final word skip and regressions out/in reflect the proportion of trials where these occurred. Figures are mean values, with standard deviations in round brackets and 95% CIs in square brackets. Skipping rate = % of words not fixated during first-pass reading; First pass RT = total time spent reading a word or phrase before exiting to the right or left; Total RT = total time spent reading a word or phrase during the trial (first-pass RT + any rereading); Regression path = total time spent reading a word or phrase and any preceding material before exiting to the right; Regressions in/out = proportion of trials where a regression was made into or out of a specified area.

figurative meaning of an idiom), but they made no improvement to any model and hence were left out. For final nouns, models for duration measures only considered items that were not skipped (hence would have durations of 0). In all models we applied a Bonferroni correction since multiple correlated measures were being tested, following the recommendations of von der Malsburg and Angele (2017). In practice, this meant adopting a new threshold for significance obtained by dividing the usual value of  $p = .05$  by the number of measures being tested. For whole phrases (five measures) this gave an adjusted threshold of  $p = .01$ ; for final nouns (five measures)  $p = .01$ ; and for the post-phrase region (four measures)  $p = .125$ . For simplicity, this means that in the results that follow a  $p$ -value of  $< .01$  is taken to represent a true difference.

For the whole phrase, idioms were read no more quickly than literal control phrases in first pass reading time ( $\beta = -0.06, t = -2.04, p = .045$ ), total reading time ( $\beta = -0.07, t = -2.54, p = .014$ ) or regression path duration ( $\beta = -0.05, t = -1.60, p = .120$ ), and there were no differences between idioms and literal controls for regressions out of the phrase ( $\beta = -0.02, z = -0.11, p = .911$ ) or regressions back in from later in the sentence ( $\beta = -0.26, z = -1.30, p = .195$ ). Compared with literal control phrases, variants were read no more slowly during first pass reading ( $\beta = -0.01, t = -0.32, p = .749$ ) and showed no difference in regression path durations ( $\beta = 0.06, t = 1.93, p = .063$ ), but had significantly longer total reading times ( $\beta = 0.11, t = 3.00, p = .004$ ). Regressions out of or into the phrase were no less likely in variants than literal controls (regressions out:  $\beta = 0.36, z = 1.96, p = .050$ ; regressions in:  $\beta = 0.34, z = 1.47, p = .141$ ). Pairwise comparison of idioms and variant phrases showed no difference in first-pass reading ( $\beta = -0.05, t = -1.48, p = .450$ ), regressions out ( $\beta = -0.39, z = -1.99, p = .137$ ), or regressions in ( $\beta = -0.60, z = -2.52, p = .034$ ), but total reading time ( $\beta = -0.18, t = -5.34, p < .001$ ) and regression path duration ( $\beta = -0.11, t = -3.60, p = .003$ ) were both shorter for idioms.

For final nouns, no measures showed a significant difference for idioms compared with literal control phrases: skipping rate ( $\beta = 0.14, z = 0.74, p = .462$ ); first pass reading times ( $\beta = -0.01, t = -0.32, p = .750$ ); total reading times ( $\beta = -0.04, t = -1.55, p = .125$ ); regression path durations ( $\beta = -0.07, t = -1.65, p = .108$ ); and regressions out ( $\beta = -0.19, z = -0.90, p = .368$ ) Likelihood of skipping the final word was no different in variants than literal controls ( $\beta = -0.26, z = -1.04, p = .301$ ), nor was the likelihood of a regression from the final noun ( $\beta = 0.23, z = 1.15, p = .251$ ). Neither first pass reading times for final nouns ( $\beta = 0.05, t = 1.83, p = .074$ ) nor regression path durations ( $\beta = 0.05, t = 1.33, p = .192$ ) were different for variants, but total reading times were significantly longer ( $\beta = 0.11, t = 2.67, p = .010$ ). Comparison of idiom and variant conditions showed



that neither skipping ( $\beta = 0.40, z = 1.86, p = .189$ ) nor regressions out ( $\beta = -0.43, z = -2.12, p = .102$ ) were more likely in idioms. First pass reading time was also no different ( $\beta = -0.06, t = 2.49, p = .055$ ), but both total reading time ( $\beta = -0.15, t = -4.27, p < .001$ ) and regression path duration ( $\beta = -0.12, t = -3.51, p = .005$ ) were significantly shorter for idioms.

In the post-phrase region, idiom sentences were no different in first pass reading ( $\beta = 0.03, t = 0.89, p = .384$ ), total reading time ( $\beta = -0.03, t = -0.97, p = .330$ ) or regression path duration ( $\beta = -0.04, t = -1.22, p = .230$ ) compared with literal controls, and regressions out of the post-phrase region were no more likely in idioms ( $\beta = -0.17, z = -0.97, p = .332$ ). Variant sentences were no different to literal controls in first pass reading ( $\beta = -0.05, t = -1.05, p = .301$ ), total reading times ( $\beta = 0.03, t = 1.12, p = .269$ ) or regression path durations ( $\beta = 0.06, t = 1.96, p = .055$ ), but regressions out of the post-phrase region were significantly more likely in variants than literal phrases ( $\beta = 0.49, z = 2.64, p = .008$ ). Comparing idioms and variants directly, there was no difference in first pass reading time ( $\beta = 0.08, t = 1.69, p = .298$ ) or total reading time ( $\beta = -0.06, t = -2.99, p = .017$ ), but regression path durations were significantly shorter in idioms ( $\beta = -0.10, t = -3.85, p = .002$ ) and regressions out of the region more likely for the variants ( $\beta = -0.67, z = -3.83, p < .001$ ).

The results overall show limited differences between conditions in early measures. Crucially, this means that variant phrases were not read for longer than idioms or literal phrases on first reading. This suggests that readers may simply have treated them as literal (albeit sometimes slightly unusual) phrases on first encounter. In later measures differences emerged whereby readers spent longer on variant phrases, compared with both idioms and literal controls. Specifically, participants spent longer reading and rereading the whole phrase and final noun of critical phrases in variants than in other conditions.<sup>5</sup> Readers also spent longer overall reading the rest of the sentence in variant conditions, compared with idioms, and regressions out of the post-phrase region were significantly more likely in variants than in either other condition. The main patterns are shown in Fig. 1 (predicted effects extracted from models described above).

### Idiom variables

We chose three measures for follow-up analysis, to reflect both early and late reading patterns: whole phrase first pass and total reading time, and post-phrase regression path duration. We added each of familiarity, decomposability and

literal plausibility (all centred to help with model fitting) to each model, first as a fixed effect then as an interaction with condition, and used log-likelihood tests to assess any improvement. As before, since we are comparing multiple measures, we applied a Bonferroni correction to reduce the likelihood of Type 1 errors, hence used an adjusted  $p$  value of  $.05/3 = .017$  (rounded up) as the threshold for significance. For brevity, we report only measures where a significant improvement was observed.

Neither familiarity nor literal plausibility made any improvement to any of the models, either as a fixed effect or interaction with condition. Decomposability improved the model as a fixed effect for post-phrase regression path duration,  $\chi^2(1) = 5.76, p = .016$ , with no further improvement as an interaction with condition. Figure 2 (left panel) shows that higher decomposability ratings led to less time overall spent reading the post-phrase context portion of the sentence (including time spent rereading earlier parts of the sentence) across all conditions ( $\beta = -0.03, t = -2.52, p = .016$ ).

### Cognitive variables

We next assessed the contribution of cognitive variables (centred). In each case, we added each variable to the original (condition only) model, first as a fixed effect then as an interaction with condition, and used log-likelihood tests to assess any improvement. As above, a corrected threshold of  $p = .017$  was used to account for the use of multiple measures.

Neither inhibitory control nor working memory made any improvement to any model. Processing speed significantly improved the model for phrase total reading time as an interaction with condition,  $\chi^2(2) = 12.02, p = .002$ . Figure 2 (right panel) suggests that processing speed may have operated differently for variants, compared with idioms and literal controls.

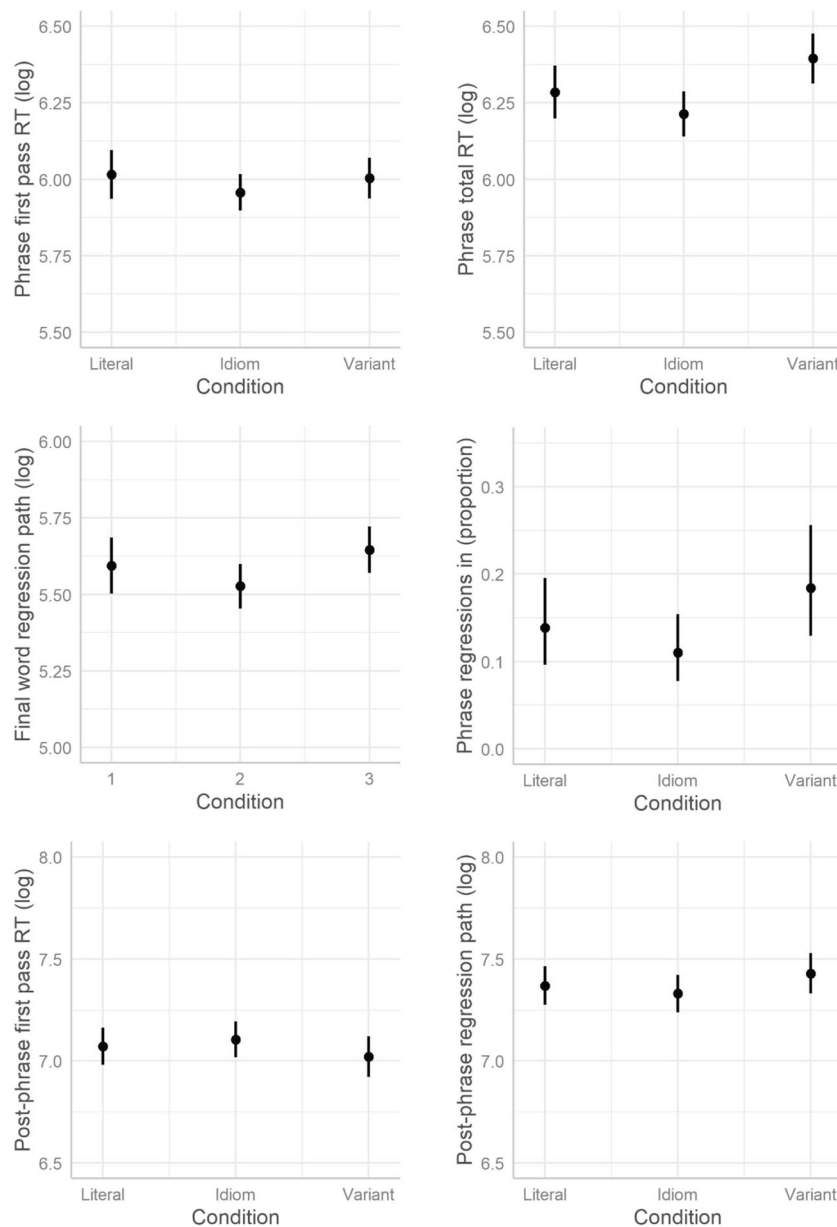
### Three-way interactions for idiom and cognitive variables

Finally, for models where idiom variables were significant, we added in cognitive variables to check for any three-way interactions. The model for post-phrase regression path duration (including a fixed effect of decomposability) was not improved by the addition of any of the cognitive variables (as fixed effects or as interactions with condition, or as part of a three-way interactions with condition and decomposability).

### Cross-modal priming analysis

Table 4 summarizes cross-modal priming data for accuracy and response times (RTs) to the figuratively related words.

<sup>5</sup> Since the final noun is included in calculation of phrase total reading time, these measures are not independent.



**Fig. 1** Predicted effects for phrase, final noun and post-phrase reading for literal, idiom, and variant phrases. Error bars show 95% confidence intervals

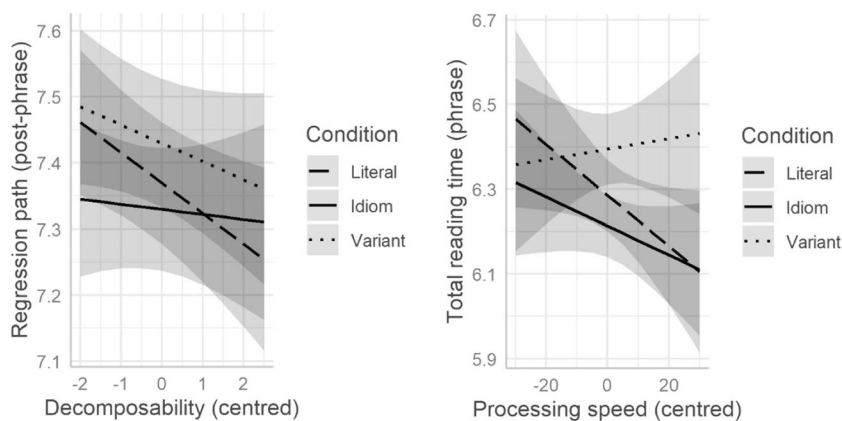
Accuracy was at ceiling across all three conditions, so we focused analysis on RTs. We first removed any incorrect responses, then removed any RTs longer than 800 ms, leading to the removal of 8% of the data and leaving 2,210 trials for analysis.

RTs were log-transformed to reduce skewing. We constructed a linear mixed-effects model with the fixed effect of condition and covariates of length and frequency (of the target word), degree of relatedness between the idiom and its target word and trial order. We included random intercepts for subject and item, by-subject random slopes for the effects of condition and relatedness, and by-item

random slopes for the effect of condition. No differences were observed between literal controls and idioms ( $\beta = 0.01, t = -1.57, p = .127$ ) or literal controls and variants ( $\beta = 0.00, t = 0.21, p = .838$ ). Neither substitutability rating nor semantic similarity between idioms and variants made a significant improvement to the model so were excluded from all further analysis.

**Idiom variables**

We next considered the same idiom variables as in the eye-tracking analysis. Neither familiarity nor literal plausibility



**Fig. 2** Left panel: Effects of decomposability on post-phrase regression path duration. Right panel: Effects of processing speed on phrase total reading times

**Table 4** Summary of cross-modal priming data for idioms, variants and literal control phrases. Response times (RTs) are in milliseconds and reflect correct answers and answers <800 ms only

Cross-modal priming	Phrase		
	Literal	Idiom	Variant
Accuracy (%)	0.99 (0.12) [0.97, 0.99]	0.99 (0.06) [0.98, 0.99]	0.99 (0.10) [0.98, 0.99]
RT	532.8 (88.8) [526, 540]	525.8 (87.0) [519, 533]	532.9 (89.4) [526, 540]

Figures are mean values, with standard deviation in round brackets and 95% CIs in square brackets.

made any improvement, but decomposability improved the model as a fixed effect,  $\chi^2(1) = 4.88$ ,  $p = .027$ . Here, higher decomposability sped up response times for all phrase types ( $\beta = -0.01$ ,  $t = -2.51$ ,  $p = .013$ ).

### Cognitive variables

We next added the cognitive variables one by one. Neither inhibitory control nor working memory made any improvements. Processing speed made an improvement as a fixed effect,  $\chi^2(1) = 6.55$ ,  $p = .011$ , but no further improvement as an interaction with condition. Faster processing speed sped up response times across all conditions ( $\beta = -0.01$ ,  $t = -3.65$ ,  $p = .011$ ).

### Three-way interactions for idiom and cognitive variables

We finally checked for any significant three-way interactions among the idiom and cognitive variables. The inclusion of a three-way interaction for condition, decomposability, and processing speed significantly improved the model,  $\chi^2(8) = 23.48$ ,  $p = .003$ , compared with the inclusion of decomposability only. Figure 3 suggests that despite this interaction, differences between conditions were minor here.

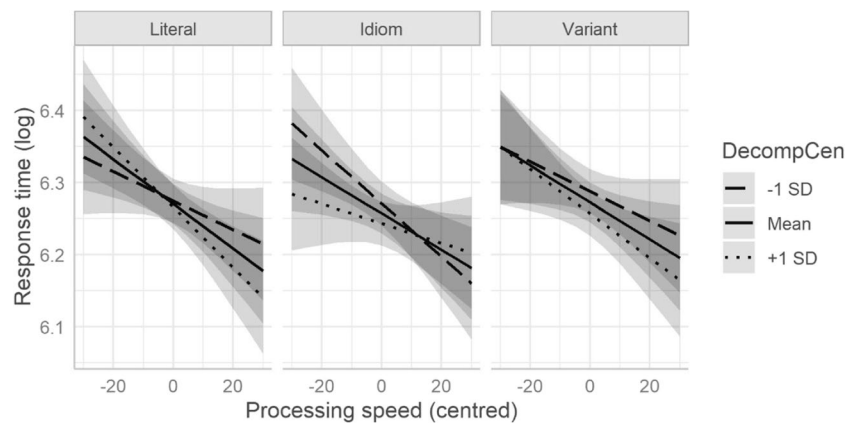
## Discussion

We examined the way in which idiom variation is processed, focusing on the effects of properties of the idioms themselves and individual cognitive differences, and comparing two different methodologies that have been widely employed in the literature—eye tracking while reading and cross-modal priming.

### Eye tracking while reading

In the eye-tracking experiment, the widely reported idiom advantage was not observed in either early or later measures (idioms were not read any more quickly than literal controls).<sup>6</sup> Of primary interest in this study was the comparison with variants: idioms modified to retain the same underlying metaphorical idea but presented in an unfamiliar lexical form. On first encounter, variants were read no differently to either idioms or literal control phrases, suggesting that readers were happy to interpret these at face value as literal

<sup>6</sup> Although note that previous studies employing eye tracking usually do not apply the more conservative approach to analysis employed here. With a less conservative approach, the difference in total reading time between idioms and literal phrases ( $p = .014$ ) would be seen as significant.



**Fig. 3** Three-way interactions between condition, decomposability, and processing speed, for log-transformed response times

statements. However, in later measures—reflecting overall time reading and rereading the critical phrase and time spent reading the rest of the sentence—a clear difference for variants emerged. The nature of the stimuli in the present study, where disambiguating context followed the target phrase, is critical here. Whilst readers saw nothing amiss with the variants on first encounter, once they had proceeded to a point where the meaning of the phrase needed to be integrated into the meaning of the whole sentence, additional processing for variants was required in order to make sense of what they had read. Two possibilities for this additional processing are (1) that the initial, literal parse of the variants was not sufficient to derive any sense from the sentence, hence a reanalysis was required; or (2) that readers were able to recognize and reconfigure the variants as modifications of known phrases, but the surprisal/novelty of these required additional processing time to resolve.

The first possibility is consistent with the studies discussed previously, where modification (especially lexical modification) generally causes difficulty and/or reduces the likelihood of idioms being recognized (Geeraert et al., 2017; Hamblin & Gibbs, 1999; Holsinger, 2013; McGlone et al., 1994; Smolka & Eulitz, 2020). Importantly, when studies have directly investigated online processing of idiomatic variation, evidence of slowed processing has tended to be most prominent in later measures (Haeuser et al., 2020; Kyriacou et al., 2020, 2021), as in the present results. Titone et al. (2019) made a distinction between earlier, “retrieval-based” and later “compositional” stages, which may further support the idea that the additional processing required for variants only emerged when more careful compositional analysis was required to make sense of the sentence as a whole. For variants, with nothing to “retrieve” on first encounter, participants had no option but to parse the phrase as literal, then reconsider this when the context demonstrated that this was problematic.

A second possibility is that readers had relatively little difficulty in recognizing variants and reconfiguring them, but the lower likelihood of the variant nouns following the verb may be reflected in the longer overall processing times. Clearly, idioms, as formulaic phrases, have higher transitional probabilities than variants (the likelihood, based on corpus frequency, that a given noun will follow a given verb; i.e. *play with* is much more likely to be followed by *fire* than *acid*), and various studies have demonstrated the effects of this variable in the eye-tracking record (Frisson et al., 2005; McDonald & Shillcock, 2003a, 2003b). Of note, Frisson et al. (2005) found effects of this in early measures in their data (as opposed to the exclusively later effects seen for our data), and also showed that transitional probability could not reliably be isolated from more general effects of predictability (e.g., from the sentence context). For these reasons, this explanation is perhaps less convincing for our results.

### Cross-modal priming

One aim of this study was to explicitly compare results from two widely used but rarely combined methodologies. Whilst eye tracking allows us to infer processing effects from reading patterns (but crucially, does not tell us directly how readers interpreted any given word or phrase), cross-modal priming offers a more direct measure of meaning activation. Given how widely this has been used to demonstrate the activation of idiom meanings in previous studies, it was surprising that we saw no “idiom superiority effect” in the present study. Since the magnitude of this effect does vary widely (from around 20 ms in Findlay & Carrol, 2019, to over 50 ms in Cacciari & Tabossi, 1988), it is possible that more studies may have found null results here, hence the near universality of this effect in the literature may reflect the so-called file drawer problem of unpublished null results (Rosenthal, 1979) and not necessarily give a true picture.

We included the degree of relatedness between idioms and their target words in the analysis to control for any differences here, hence the lack of an effect was not simply due to the target words being too far removed from the intended figurative meanings of the idioms.

Importantly for this study, no difference at all was observed between literal and variant conditions, confirming at least that variant phrases generated no consideration of figurative meaning in the timeframe employed here (250 ms following the offset of the phrase). Blasko and Connine (1993) found (mixed) evidence that even low familiar metaphors showed some degree of priming for figuratively related targets at 750 ms post-offset, so it may be that the time course of our study was not sufficient to capture any effects for variants. Combined with the lack of early effects from the eye-tracking study, it is likely that participants simply treated the variants as literal phrases in the first instance. Any attempt to reconfigure/reinterpret these was time-consuming, as observed in the eye-tracking data, and clearly took longer than the 250 ms allotted in the cross-modal priming study.

### Effects of idiom and cognitive variables

Effects of other variables were limited in both the eye-tracking and cross-modal priming results.<sup>7</sup> Surprisingly, given its overall importance in idiom processing, familiarity had no effects on any measures for any of the phrase types in either study. We hypothesized that this may have an effect in one of two ways: either more familiar idioms would be more entrenched and hence may be less amenable to variation (cf. Haeuser et al., 2020); alternatively, more familiar idioms may be better known and hence easier to reconfigure. Since neither pattern emerged, there is little more to conclude here, although the overall high levels of familiarity may mean that any potential for variation was masked here. Similarly, literal plausibility had no effect on any phrase type, but decomposability did demonstrate an effect in both studies. In the eye-tracking data, it affected post-phrase regression path durations (time spent reading the post-phrase region, plus any time spent rereading the phrase itself, before the trial ended), with more decomposable phrases from all conditions showing shorter regression paths. In the cross-modal priming study, decomposability sped up response times for all phrase types. In both cases, this suggests that when figurative and literal readings were more closely aligned (as in examples like *lift your spirits / lift your feelings*), processing

was overall easier (for both idioms and variants) than when the figurative and literal readings were further apart (e.g., *walk on air / walk on wind*).

The cognitive variables explored here (working memory, processing speed and inhibitory control) likewise had limited effects on processing in the two studies. Given the somewhat mixed picture presented in the previous literature, this is not altogether surprising. Studies of metaphor processing have generally found an effect of variables such as working memory and inhibitory control, but idioms have sometimes shown these effects (e.g., Cacciari et al., 2018, for cross-modal priming) and sometimes not (e.g., Columbus et al., 2015, for eye-tracking). If variants were treated as novel metaphorical expressions, we would have expected them to show results more akin to those found for metaphors, but no such patterns were found for any of the eye-tracking measures observed here. The two variables that we most expected to have an effect based on previous studies were working memory (Blasko, 1999; Carriedo et al., 2016; Chiappe & Chiappe, 2007; Columbus et al., 2015; Kazmerski et al., 2003; Olkonemi et al., 2016; Pierce et al., 2010) and inhibitory control (Cacciari et al., 2018; Papagno et al., 2003). Neither had an effect on idioms or variants in either study, which further suggests that variants were treated more like literal phrases on first encounter (i.e., no competition existed between figurative and literal meanings; hence, neither working memory nor inhibitory control came into play to help readers “commit” to one interpretation or the other). The one variable that did show an effect was processing speed, where faster processors read idioms and literal controls, but not variants, more quickly than slower processors (Fig. 2, right panel). As with the overall patterns, this effect was only apparent in later measures (total reading time). This may support the overall interpretation that variants were treated as literal on first encounter, but required significant later analysis (for all readers, regardless of natural processing speed) once the remaining sentence context had been encountered.

Similarly, in the cross-modal priming study, processing speed decreased response times across the board, but neither working memory nor inhibitory control made any improvement to the analysis. As argued above, this strengthens the assumption that variants were treated as literal phrases on first encounter, although effects of these variables may become apparent at a longer time course (see, e.g., the differential effects observed at different presentation speeds in Libben & Titone, 2014).

### Overall summary and conclusions

Existing literature on idiom variation shows a broad distinction between studies of interpretation, where idioms can still be successfully understood even when modified in various ways, and studies of online processing, where a clear cost is incurred by creative manipulation. In the present findings,

<sup>7</sup> We note here that the lack of clear and consistent idiom superiority effects, and the lack of effects for cognitive and idiom variable, may be a power issue, and future studies might usefully test much larger samples, especially when individual differences are being investigated.

that cost showed up in later measures of reading (generally thought to reflect meaning integration processes), but not in earlier measures. Of importance here is the fact that a comparison of early versus late measures, as we have adopted in this study, is a simplistic and possibly even misleading distinction (Clifton et al., 2007). For one, the measures we use are not independent of one another, hence a “late” measure like total reading time by definition also includes “early” measures such as first-pass reading time. This makes it much harder to neatly separate these into early (recognition) and later (meaning selection/integration) processes, and Vasishth et al. (2012) argue that critical effects may well be subject to a “lag,” whereby they start early but are only observable later in the processing record. Whilst it is therefore not possible to conclusively say that variants were initially treated as literal phrases, with subsequent processing only required at a later stage, it is clear that variants posed little difficulty for readers during the initial parse of the sentence but did induce subsequently more rereading than either literal controls or known idioms. In comparison, Carrol and Littlemore (2020) performed a similar study with entirely unfamiliar idioms and found significantly slower reading (compared with a literal paraphrase of the same meaning) in early as well as late measures (all at the  $p < .001$  level, so these results would have been significant even with the more conservative analysis adopted in the present study). Tentatively, then, we conclude from the present study that the challenge posed by variants was one that primarily, although possibly not exclusively, manifested itself in the process of making sense of the phrase in context, rather than during the initial stages of lexical recognition and access.

With this in mind, we further assume that the additional total reading time observed for variants is a result of the need to reconsider the variants and assign an alternative figurative meaning, but may also reflect surprisal at an unfamiliar version of a known phrase. On balance, and for the reasons outlined earlier in this section, we favour the former interpretation of our results here. Importantly, there is little evidence that variants were treated as novel metaphors, given the lack of an effect of variables such as working memory, which have shown such consistent effects in the metaphor literature. In cross-modal priming, which directly tests the activation of meaning, no consideration of the figurative meaning of variants was observed, although the lack of an idiom advantage for canonical phrases means that this is unsurprising in the present study, and far from conclusive. The lack of clear and consistent effects for the cognitive variables studied here leaves the question of their contribution open, since previous studies have also found mixed results. The use of two complementary approaches here allowed us to directly compare methods, and both produced results that were less clear in terms of an idiom advantage than is generally reported in the literature.

## Appendix: Stimulus items used in both studies

Idiom	Variant	Literal control
<i>bite the dust</i>	<i>bite the dirt</i>	<i>bite the meat</i>
<i>blow your cover</i>	<i>blow your shelter</i>	<i>blow your nose</i>
<i>break the ice</i>	<i>break the shell</i>	<i>break the bowl</i>
<i>burn your bridges</i>	<i>burn your routes</i>	<i>burn your dinner</i>
<i>bury the hatchet</i>	<i>bury the cleaver</i>	<i>bury the embers</i>
<i>call the shots</i>	<i>call the moves</i>	<i>call the kids</i>
<i>change your tune</i>	<i>change your song</i>	<i>change your job</i>
<i>clear your name</i>	<i>clear your image</i>	<i>clear your books</i>
<i>cover your tracks</i>	<i>cover your steps</i>	<i>cover your plate</i>
<i>crack the whip</i>	<i>crack the belt</i>	<i>crack the pane</i>
<i>cut your losses</i>	<i>cut your debts</i>	<i>cut your grass</i>
<i>drag your feet</i>	<i>drag your legs</i>	<i>drag your bags</i>
<i>draw a blank</i>	<i>draw a zero</i>	<i>draw a clown</i>
<i>drop the ball</i>	<i>drop the baby</i>	<i>drop the glass</i>
<i>drown your sorrows</i>	<i>drown your sadness</i>	<i>drown your lilies</i>
<i>eat your words</i>	<i>eat your ideas</i>	<i>eat your food</i>
<i>face the music</i>	<i>face the songs</i>	<i>face the front</i>
<i>fall from grace</i>	<i>fall from esteem</i>	<i>fall from height</i>
<i>fan the flames</i>	<i>fan the blaze</i>	<i>fan the queen</i>
<i>get the picture</i>	<i>get the image</i>	<i>get the letter</i>
<i>grease the wheels</i>	<i>grease the chain</i>	<i>grease the bike</i>
<i>hit the road</i>	<i>hit the street</i>	<i>hit the door</i>
<i>hold the fort</i>	<i>hold the camp</i>	<i>hold the spoon</i>
<i>jump the gun</i>	<i>jump the shot</i>	<i>jump the fence</i>
<i>learn the ropes</i>	<i>learn the moves</i>	<i>learn the rules</i>
<i>lick your wounds</i>	<i>lick your injuries</i>	<i>lick your finger</i>
<i>lift your spirits</i>	<i>lift your feelings</i>	<i>lift your weights</i>
<i>lose the thread</i>	<i>lose the strand</i>	<i>lose the fight</i>
<i>meet your match</i>	<i>meet your equal</i>	<i>meet your uncle</i>
<i>miss the boat</i>	<i>miss the ship</i>	<i>miss the play</i>
<i>over the moon</i>	<i>over the stars</i>	<i>over the fence</i>
<i>pass the torch</i>	<i>pass the lamp</i>	<i>pass the dish</i>
<i>play with fire</i>	<i>play with acid</i>	<i>play with toys</i>
<i>pull some strings</i>	<i>pull some wires</i>	<i>pull some blankets</i>
<i>rock the boat</i>	<i>rock the ship</i>	<i>rock the chair</i>
<i>roll with the punches</i>	<i>roll with the blows</i>	<i>roll with the waves</i>
<i>save your skin</i>	<i>save your bones</i>	<i>save your cash</i>
<i>scratch the surface</i>	<i>scratch the shell</i>	<i>scratch the mirror</i>
<i>shake a leg</i>	<i>shake a foot</i>	<i>shake a tree</i>
<i>sing your praises</i>	<i>sing your merits</i>	<i>sing your tunes</i>
<i>sit on the fence</i>	<i>sit on the wall</i>	<i>sit on the grass</i>
<i>smell a rat</i>	<i>smell a toad</i>	<i>smell a rose</i>
<i>speak your mind</i>	<i>speak your views</i>	<i>speak your lines</i>
<i>spill the beans</i>	<i>spill the seeds</i>	<i>spill the juice</i>
<i>steal the show</i>	<i>steal the play</i>	<i>steal the gold</i>
<i>swallow your pride</i>	<i>swallow your dignity</i>	<i>swallow your tablet</i>
<i>throw in the towel</i>	<i>throw in the cloth</i>	<i>throw in the coins</i>

<i>turn the tables</i>	<i>turn the boards</i>	<i>turn the pages</i>
<i>under the weather</i>	<i>under the elements</i>	<i>under the bridge</i>
<i>walk on air</i>	<i>walk on wind</i>	<i>walk on stage</i>
<i>waste your breath</i>	<i>waste your oxygen</i>	<i>waste your petrol</i>

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13421-023-01463-x>.

**Open practices statement** All data and materials are available (<https://osf.io/g54mb/>), and this study was not preregistered.

**Authors' contributions** The first author designed and built the studies, prepared materials and collected and analyzed the data. The second author advised on appropriate cognitive tests to include and contributed to material refinement. Both authors contributed to the writing and redrafting of the paper.

**Funding** No funding was provided for this study.

**Data availability** All data and materials are available (<https://osf.io/g54mb/>).

**Code availability** Not applicable.

## Declarations

**Conflicts of interest** No conflicts of interest exist.

**Ethics approval** The study was approved by the College of Arts and Law Ethics Committee at the University of Birmingham.

**Consent to participate** All participants provided signed consent to participate and were informed of their right to withdraw at any point.

**Consent for publication** All participants provided signed consent for the anonymized data to be used in publications.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Memory & Cognition*, 24(4), 477–492. <https://doi.org/10.3758/BF03200936>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. <https://doi.org/10.1016/j.jml.2009.09.005>

- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248. [10.1111/j.1467-9280.2008.02075.x](https://doi.org/10.1111/j.1467-9280.2008.02075.x)
- Blasko, D. G. (1999). Only the tip of the iceberg: Who understands what about metaphor? *Journal of Pragmatics*, 31(12), 1675–1683. [https://doi.org/10.1016/S0378-2166\(99\)00009-0](https://doi.org/10.1016/S0378-2166(99)00009-0)
- Blasko, D., & Briihl, D. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol*, 12(4), 261–285. [https://doi.org/10.1207/s15327868ms1204\\_4](https://doi.org/10.1207/s15327868ms1204_4)
- Blasko, D., & Connine, C. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(2), 295–308. <https://doi.org/10.1037/0278-7393.19.2.295>
- Bulkes, N., & Tanner, D. (2017). “Going to town”: Large-scale norming and statistical analysis of 870 American English idioms. *Behavior Research Methods*, 49(2), 772–783. <https://doi.org/10.3758/s13428-016-0747-8>
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27(6), 668–683. [https://doi.org/10.1016/0749-596X\(88\)90014-9](https://doi.org/10.1016/0749-596X(88)90014-9)
- Cacciari, C., Corradini, P., & Ferlazzo, F. (2018). Cognitive and personality components underlying spoken idiom comprehension in context. An exploratory study. *Frontiers in Psychology*, 9, 659. <https://doi.org/10.3389/fpsyg.2018.00659>
- Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: evidence for a new hybrid view. *Metaphor and Symbol*, 22(1), 79–108. [https://doi.org/10.1207/s15327868ms2201\\_3](https://doi.org/10.1207/s15327868ms2201_3)
- Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLOS ONE*, 11(3), e0150289. <https://doi.org/10.1371/journal.pone.0150289>
- Carrol, G. (2021). Psycholinguistic approaches to figuration. In A. Soares da Silva (Ed.), *Figurative language—Intersubjectivity and usage* (pp. 307–338). John Benjamins.
- Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17(4), 784–797. <https://doi.org/10.1017/S1366728913000795>
- Carrol, G., & Conklin, K. (2017). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea “has legs”. *Bilingualism: Language and Cognition*, 20(2), 299–317. <https://doi.org/10.1017/S1366728915000103>
- Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 63(1), 95–122. <https://doi.org/10.1177/0023830918823230>
- Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of L1 on the processing of idioms in L2. *Studies in Second Language Acquisition*, 38(3), 403–443. <https://doi.org/10.1017/s0272263115000492>
- Carrol, G., & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, 57(7), 609–626. <https://doi.org/10.1080/0163853X.2020.1729041>
- Carrol, G., Littlemore, J., & Dowens, M. G. (2018). Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua*, 204, 21–44. <https://doi.org/10.1016/j.lingua.2017.11.001>
- Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, 56(2), 172–188. <https://doi.org/10.1016/j.jml.2006.11.006>
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. V. Gompel, M. Fisher, W.

- Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier.
- Columbus, G., Sheikh, N. A., Cote-Lecaldare, M., Hauser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience*, 8, 1057. <https://doi.org/10.3389/fnhum.2014.01057>
- Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.
- Cronk, B., & Schweigert, W. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. *Applied Psycholinguistics*, 13, 131–146. <https://doi.org/10.1017/S0142716400005531>
- Duffley, P. J. (2013). How creativity strains conventionality in the use of idiomatic expressions. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), *Language and the creative mind* (pp. 49–61). CSLI Publications.
- Fellbaum, C. (2019). How flexible are idioms? A corpus-based study. *Linguistics*, 57(4), 735–767. <https://doi.org/10.1515/ling-2019-0015>
- Feyaerts, K. (2006). Towards a dynamic account of phraseological meaning: Creative variation in headlines and conversational humour. *International Journal of English Studies*, 6(1), 57–84.
- Findlay, H., & Carrol, G. (2019). Contributions of semantic richness to the processing of idioms. *The Mental Lexicon*, 13(3), 311–333. <https://doi.org/10.1075/ml.18014.fin>
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 862–877. <https://doi.org/10.1037/0278-7393.31.5.862>
- Geeraert, K., Newman, J., & Baayen, R. H. (2017). Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science*, 9(3), 653–669. <https://doi.org/10.1111/tops.12263>
- Gibbs, R. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8, 149–156. <https://doi.org/10.3758/BF03213418>
- Glucksberg, S. (2001). *Understanding figurative language: From metaphor to idioms*. Oxford University Press.
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.
- Haeuser, K. I., Baum, S., & Titone, D. (2020). Effects of aging and noncanonical form presentation on idiom processing: Evidence from eye tracking. *Applied Psycholinguistics*, 42(1), 101–127. <https://doi.org/10.1017/S0142716420000612>
- Hamblin, J., & Gibbs, R. (1999). Why you can't kick the bucket as you slowly die: Verbs in idiom comprehension. *Journal of Psycholinguistic Research*, 28(1), 25–39. <https://doi.org/10.1023/A:1023235403250>
- Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech*, 56(3), 373–394. <https://doi.org/10.1177/0023830913484899>
- Hovhannisyian, A., & Mkrtchyan, A. (2014). Transformation of idioms and transparency. *Journal of Modern Education Review*, 4(1), 29–34.
- Inhoff, A. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612–624. [https://doi.org/10.1016/S0022-5371\(84\)90382-7](https://doi.org/10.1016/S0022-5371(84)90382-7)
- Inhoff, A., Lima, S., & Carroll, P. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, 12(6), 558–567. <https://doi.org/10.3758/BF03213344>
- Kazmerski, V. A., Blasko, D. G., & Dessalegn, B. G. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory & Cognition*, 31(5), 673–689. <https://doi.org/10.3758/BF03196107>
- Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, 34(1), 89–109. <https://doi.org/10.1006/jmla.1995.1005>
- 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>
- Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, 63(2), 404–435. <https://doi.org/10.1177/0023830919847691>
- Kyriacou, M., Conklin, K., & Thompson, D. (2021). When the Idiom advantage comes up short: Eye-tracking canonical and modified idioms. *Frontiers in Psychology*, 12, 675046. <https://doi.org/10.3389/fpsyg.2021.675046>
- Langlotz, A. (2006a). *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. John Benjamins.
- Langlotz, A. (2006b). Occasional adnominal idiom modification—A cognitive linguistic approach. *International Journal of English Studies*, 6(1), 85–108.
- Libben, M., & Titone, D. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, 36(6), 1103–1121. <https://doi.org/10.3758/MC.36.6.1103>
- Mancuso, A., Elia, A., Laudanna, A., & Vietri, S. (2020). The role of syntactic variability and literal interpretation plausibility in idiom comprehension. *Journal of Psycholinguistic Research*, 49, 99–124. <https://doi.org/10.1007/s10936-019-09673-8>
- McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648–652. <https://doi.org/10.1046/j.0956-7976.2003.psci.1480.x>
- McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735–1751. [https://doi.org/10.1016/S0042-6989\(03\)00237-2](https://doi.org/10.1016/S0042-6989(03)00237-2)
- McGlone, M., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, 17, 167–190. <https://doi.org/10.1080/01638539409544865>
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Clarendon Press.
- Mueller, R. A., & Gibbs, R. W. (1987). Processing idioms with multiple meanings. *Journal of Psycholinguistic Research*, 16(1), 63–81.
- Oliveri, M., Papagno, C., & Romero, L. (2004). Left but not right temporal lobe involvement in opaque idiom comprehension: A repetitive transcranial stimulation study. *Journal of Cognitive Neuroscience*, 16(5), 848–855. <https://doi.org/10.1162/089892904970717>
- Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 433–450. <https://doi.org/10.1037/xlm0000176>
- Papagno, C., & Caporali, A. (2007). Testing idiom comprehension in aphasic patients: The modality and the type of idiom effects. *Brain & Language*, 100(2), 208–220. <https://doi.org/10.1016/j.bandl.2006.01.002>
- Papagno, C., & Genoni, A. (2004). The role of syntactic competence in idiom comprehension: A study on aphasic patients. *Journal of Neurolinguistics*, 17(5), 371–382. <https://doi.org/10.1016/j.jneuroling.2003.11.002>
- Papagno, C., Lucchelli, F., Muggia, S., & Rizzo, S. (2003). Idiom comprehension in Alzheimer's disease: The role of the central executive. *Brain*, 126(11), 2419–2430. <https://doi.org/10.1093/brain/awg243>
- Papagno, C., Tabossi, P., Colombo, M., & Zampetti, P. (2004). Idiom comprehension in aphasic patients. *Brain & Language*, 89(1), 226–234. [https://doi.org/10.1016/S0093-934X\(03\)00398-5](https://doi.org/10.1016/S0093-934X(03)00398-5)



- Papagno, C., Curti, R., Rizzo, S., Crippa, F., & Colombo, M. R. (2006). Is the right hemisphere involved in idiom comprehension? A neuropsychological study. *Neuropsychology*, 20(5), 598–606. <https://doi.org/10.1037/0894-4105.20.5.598>
- Paterson, K., Liversedge, S., & Underwood, G. (1999). The influence of focus operators on syntactic processing of short relative clause sentences. *The Quarterly Journal of Experimental Psychology*, 52A, 717–737. <https://doi.org/10.1080/713755827>
- Pierce, R. S., MacLaren, R., & Chiappe, D. L. (2010). The role of working memory in the metaphor interference effect. *Psychonomic Bulletin & Review*, 17(3), 400–404. <https://doi.org/10.3758/PBR.17.3.400>
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55A(4), 1339–1362. <https://doi.org/10.1080/02724980244000099>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762–776. [https://doi.org/10.1162/jocn\\_a\\_00337](https://doi.org/10.1162/jocn_a_00337)
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Schweigert, W. (1986). The comprehension of familiar and less familiar idioms. *Journal of Psycholinguistic Research*, 15, 33–45. <https://doi.org/10.1007/BF01067390>
- Schweigert, W. (1991). The muddy waters of idiom comprehension. *Journal of Psycholinguistic Research*, 20(4), 305–314. <https://doi.org/10.1007/BF01074283>
- Schweigert, W., & Moates, D. (1988). Familiar idiom comprehension. *Journal of Psycholinguistic Research*, 17(4), 281–296. <https://doi.org/10.1007/BF01067198>
- Senaldi, M., Wei, J., Gullifer, J., & Titone, D. (2022). Scratching your tête over language-switched idioms: Evidence from eye-movement measures of reading. *Memory & Cognition*, 50(6), 1230–1256. <https://doi.org/10.3758/s13421-022-01334-x>
- Siyanova-Chanturia, A., & Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics*, 35(5), 549–569. <https://doi.org/10.1093/applin/amt054>
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase ‘time and again’ matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37(3), 776–784. <https://doi.org/10.1037/a0022531>
- Smolka, E., & Eulitz, C. (2020). Can you reach for the planets or grasp at the stars?—Modified noun, verb, or preposition constituents in idiom processing. In S. Schulte im Walde & E. Smolka (Eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective* (pp. 179–204). Language Science Press. <https://doi.org/10.5281/zenodo.3598566>
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–184. <https://doi.org/10.1016/j.jml.2005.11.001>
- Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, 18(03), 419–437. <https://doi.org/10.1017/s1366728914000674>
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659. [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4)
- Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523–534. [https://doi.org/10.1016/S0022-5371\(79\)90284-6](https://doi.org/10.1016/S0022-5371(79)90284-6)
- Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition*, 7(3), 159–165. <https://doi.org/10.3758/BF03197534>
- Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, 43, 110–123.
- Tabossi, P., Fanari, R., & Wolf, K. (2005). Spoken idiom recognition: Meaning retrieval and word expectancy. *Journal of Psycholinguistic Research*, 34, 465–495. <https://doi.org/10.1007/s10936-005-6204-y>
- Tabossi, P., & Zardon, F. (1993). The activation of idiomatic meaning in spoken language comprehension. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, structure and interpretation* (pp. 145–161). Erlbaum.
- Tabossi, P., & Zardon, F. (1995). The activation of idiomatic meaning. In M. Everaert, E. van den Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 273–282). Erlbaum.
- Tabossi, P., Fanari, R., & Wolf, K. (2009a). Why are idioms recognized fast? *Memory and Cognition*, 37(4), 529–540. <https://doi.org/10.3758/MC.37.4.529>
- Tabossi, P., Wolf, K., & Koterle, S. (2009b). Idiom syntax: Idiosyncratic or principled? *Journal of Memory and Language*, 61(1), 77–96.
- Titone, D., & Connine, C. (1994a). Comprehension of idiomatic expressions: Effects of predictability and literalness. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(5), 1126–1138. <https://doi.org/10.1037/0278-7393.20.5.1126>
- Titone, D., & Connine, C. (1994b). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literalness. *Metaphor and Symbolic Activity*, 9(4), 247–270.
- Titone, D., & Connine, C. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, 31, 1655–1674. [https://doi.org/10.1016/S0378-2166\(99\)00008-9](https://doi.org/10.1016/S0378-2166(99)00008-9)
- Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon*, 9(3), 473–496. <https://doi.org/10.1075/ml.9.3.05tit>
- Titone, D., Lovseth, K., Kasparian, K., & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 73(4), 216–230. <https://doi.org/10.1037/cep0000175>
- Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. B. Cieslicka (Eds.), *Bilingual figurative language processing* (pp. 171–207). Cambridge University Press.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24, 330–335. <https://doi.org/10.1044/jshr.2403.330>
- Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 42(10), 1632–1642. <https://doi.org/10.1037/xlm0000259>

- von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. <https://doi.org/10.1016/j.jml.2016.10.003>
- Vrbinc, A., & Vrbinc, M. (2011). Creative use of idioms in satirical magazines. *Jezikoslovlje*, *12*(1), 75–91.
- Wechsler, D. (2010). Wechsler Adult Intelligence Scale–Fourth UK Edition (WAIS-IV UK). Pearson Assessment.
- Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. Continuum International Publishing Group.
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.