UNIVERSITYOF BIRMINGHAM

Comprehension of different types of novel metaphors in monolinguals and multilinguals

Werkman-Horvat, Ana; Bolognesi, Marianna; Littlemore, Jeannette; Barnden, John

[Link to publication on Research at Birmingham portal](#)

Comprehension Of Different Types Of Novel Metaphors In monolInguals And multilInguals

Ana Werkmann Horvat[1] (iD), Marianna Bolognesi[2*], Jeannette Littlemore[3] and John Barnden[3]

[1]University of Osijek; [2]University of Bologna; [3]University of Birmingham
*Corresponding author. Email: m.bolognesi@unibo.it

Abstract
It has been suggested that multilingualism can lead to increased cognitive flexibility and creativity. No studies to date, however, have investigated whether this advantage leads to a greater propensity to find meaning in different kinds of novel metaphors. This article reports a self-paced reading study that focuses on whether such an increased propensity is displayed by multilingual English speakers, as opposed to monolingual English speakers. The article explores the difference between two broad types of novelty in metaphorical expressions, which are distinguished by how readily they conform to existing metaphorical schemata. The results indicate that both monolinguals and multilinguals find novel metaphors that conform readily to an existing schema easier to comprehend those that do not. They also take longer to seek meaning in metaphors that conform readily to an existing schema. Multi-linguals are more likely than monolinguals to find meaning in both types of novel metaphor. The theoretical distinction drawn between metaphors that conform readily to an existing schema and those that do not highlights the variability of meaning in novel metaphors. It also focuses attention on the different extents to which hearers seek rich meanings as opposed to less rich but more easily derived ones.

1. Introduction

This work originated in two different but interacting interests: (1) the ways in which being multilingual rather than monolingual affects one's comprehension of novel metaphor; and (2) the diverse ways in which metaphorical utterances can be novel, and how this diversity affects comprehension of the utterances.

There is evidence to suggest that being multilingual leads to increased cognitive flexibility (Bialystok, 2011), although, as we will see below, this idea has been challenged. Prompted by this, we ask whether such cognitive flexibility effects arise in novel metaphor comprehension. Specifically, we investigate whether multilinguals

more readily comprehend novel metaphorical utterances than monolinguals do. That is, we explore whether multilinguals are more inclined to see meaning in novel metaphorical utterances than monolinguals are. At the same time, we are interested in whether any such effect differs between two different types of metaphor novelty that we introduce, based partly on work by Barnden (2015), and that we call here 'high-conforming novelty' and 'low-conforming novelty'. These types of novelty reflect differing extents to which novel metaphors that conform to an existing schema.

The article proceeds as follows. The next section provides theoretical and experimental background, and motivates the particular hypotheses for experimental investigation in our study. Section 3 presents the detailed nature of the experiment. Sections 4 and 5 report and discuss the results, showing how they broadly support our hypotheses concerning the impact of metaphor novelty type on comprehension difficulty and the tendency for multilingualism to enhance comprehension for both novelty types. Finally, the Supplementary Document[1] explains why, in the experiment, we classified particular metaphorical examples in the way that we did.

2. Background, approach and hypotheses

2.1. Multilinguals, monolinguals and metaphor

One might hypothesise that multilingual individuals are more flexible than monolinguals when seeking to comprehend novel metaphor. There is a substantial body of literature indicating that multilinguals are, generally speaking, more cognitively flexible than monolinguals. Classic empirical studies show that multilinguals appear to outperform monolinguals on cognitive control tasks (Bialystok, 2001a, 2001b; Bialystok et al., 2014). Cognitive control is the ability to deal with potentially conflicting sources of information, to ignore irrelevant sources and to deliberately switch between sources. Multilinguals' superior performance suggests that learning another language renders our categorisation systems more flexible (Bialystok & Martin, 2004; Jacques & Zelazo, 2001; Martin-Rhee & Bialystok, 2008). In addition, multilinguals appear to develop stronger metalinguistic skills than monolinguals (Bialystok, 2001a; Bialystok et al., 2014), and this has a washback effect on their first language (L1) ability (Jarvis, 2003; Murphy & Pine, 2003; Yelland, Pollard, & Mercuri, 1993).

The issue of multilinguals' higher cognitive flexibility, compared to monolinguals, is not free from controversy and has, in recent years, been at the centre of a replication crisis. Recent meta-analyses of studies that claim to have found cognitive advantages for multilinguals conclude that some differences in executive control tasks may have been overstated (Lehtonen et al., 2018; Paap & Greenberg, 2013; Papageorgiou et al., 2019). Other meta-analyses suggest that such differences may be linked to the publication bias phenomenon, in which empirical studies that observe a significant difference are more likely to be published than studies that report no statistical effect (de Bruin, Bak, & Della Sala, 2015; de Bruin, Dick, & Carreiras, 2021; de Bruin, Treccani, & Della Sala, 2015a, 2015b).

It therefore remains unclear whether, to what extent, and in what ways monolinguals and multilinguals differ in their repertoire of cognitive and linguistic

---

[1]Accessible at https://osf.io/ek4q8/?view_only=faa82d8334fd478cb8f99fd15f107597.

abilities. As a step towards removing some of this lack of clarity in the field, we focus specifically on metaphor, and test the hypothesis that multilingual speakers display greater flexibility in dealing with novel metaphor, compared to monolinguals. This general hypothesis is supported by empirical literature in metaphor studies, showing, for instance, that on acquiring a second language (L2), people more frequently produce new metaphors in their L1 (Kecskés & Papp, 2000). These authors suggest that L2 acquisition engenders greater ability to construe events in different ways: it leads to a common underlying conceptual base for the two language channels, which constantly interact. Presumably, therefore, multilinguals can incorporate metaphorical associations from both languages, access them in either language and switch between them easily. Other studies have suggested strong links between divergent thinking (namely, the tendency to make fast associative connections between distant ideas aimed at exploring potential solutions; see Runco & Acar, 2012), fast automatic holistic processing and creative metaphor production and comprehension, particularly in L2s (Birdsell, 2018a, 2018b; Littlemore, 2001, 2010; Littlemore & Low, 2006a, 2006b). In addition, people are more likely to find meaning in novel metaphor when operating in their L2 than when operating in their L1 (Littlemore, 2010). These studies suggest that multilinguals will be more flexible in comprehending novel metaphor, in the sense of being more likely to find meaning in it.

It is less clear how multilingualism might affect the speed with which novel metaphorical expressions are processed. Fast metaphor meaning identification correlates with holistic processing (Littlemore, 2001), suggesting that it relies largely on loose associative networks and coarse semantic processing (Beeman, 1998). The fact that multilinguals have more elaborate associative networks than monolinguals may mean that it takes them longer to search these networks for meaning. This tentative hypothesis is indirectly supported by robust findings in lexical retrieval tasks where multilinguals tend to perform more slowly than monolinguals, and exhibit more effortful language processing (Bialystok et al., 2009).

2.2. Conventionality and two types of novelty

Numerous studies have investigated differences between novel and conventional metaphors (see below for further information on how we define these terms), comparing, for instance, the speed with which they are comprehended. The broad conclusion from such studies appears to be that novel metaphors take longer to process (Cacciari et al., 2011; Cardillo et al., 2012). Blasko & Connine (1993) and Blasko & Briihl (1997) demonstrated a clear effect of familiarity on how metaphors are comprehended (see also Bambini et al., 2019; Columbus et al., 2015; Mashal & Faust, 2009). Similarly, Bowdle & Gentner (2005) presented their career of metaphor hypothesis, whereby novel metaphors are comprehended through intricate analogy building, but conventional metaphors (at least of the A-is-B type) are comprehended through a less effortful process of categorisation.

In such studies, there are some gradations of novelty and conventionality, and these qualities are generally regarded as occupying different regions of a single, simple scale. There have, however, been few attempts to explore whether qualitatively different ways of being novel have a bearing on comprehension difficulty, and if so, whether this relates to mono/multilingualism. We mention here two studies that

have investigated phenomena that are related to these issues. First, Littlemore et al. (2018) studied the role that 'optimal innovation' plays (Giora et al., 2004) in metaphor comprehension, and found that 'optimally innovative' metaphors were more likely to be deemed to be of a higher quality than metaphors that were 'too' creative and therefore not optimally innovative. Second, Werkmann Horvat, Bolognesi, & Kohl (2021)) explored the ways in which easy-to-interpret and difficult-to-interpret metaphors were received by people with and without multilingual experience. They found that people with multilingual experience were more likely to say that the metaphor makes sense than those without multilingual experience. However, neither of these studies discussed the differences between different types of novelty in any depth. Before further addressing novelty, we must address conventionality.

2.2.1. Conventionality

Usually, a metaphor is considered to be conventional if it has an established meaning that is widely shared by L1 speakers of the language.[2] The matter can differ between different theorists, notably because of different views of what counts as metaphorical. Another complication is that the degree of familiarity a language user might have with the conventional term and its meaning(s) can vary between users and between expressions. Nevertheless, there are many expressions, such as 'at the back of one's mind', which have an established meaning that most theorists would regard as metaphorical and that would be easily discerned by L1 speakers.

Following Müller (2009), we take conventional metaphor to come in two broad subtypes, 'transparent' and 'opaque', with the latter subtype equated by Müller with the type that should truly be said to be 'dead' or 'historical' (e.g., 'broken heart'). The transparent metaphors, which are the only ones of explicit interest in our study, are those where L1 speakers are equipped to see how the metaphorical meaning fits with extant literal meanings of the words in the metaphor. In Section 3.1.1, we comment further on our particular criteria for selecting conventional metaphorical terms for our particular study.

In line with Müller's (2009) claims that conventional metaphorical meaning, if sufficiently common and standard, is easily retrievable by hearers, and in line with evidence about the speed of comprehending conventional metaphor, even when transparent and out of context (see, e.g., Holyoak & Stamenković, 2018), we assume that conventional metaphorical meanings are *usually* simply retrieved, in much the same way as literal meanings are generally assumed to be. However, Müller (2009) mentions various types of special circumstance where something more than straightforward retrieval can happen, including circumstances where the term in question is accompanied by pictures, hand gestures or other linguistic metaphorical terms using the same source subject matter.[3] However, even without a special circumstance, we should

---

[2]The notion of conventionality is rarely as precisely defined in the metaphor research literature as one might wish, although it is necessarily operationalised in particular ways in the many psycholinguistic studies that feature conventional metaphor. Our description of it incorporates the portrayal of conventional metaphor in Müller, 2009 (see especially p. 181 and elsewhere in Ch. 6), and accords with the notion of conventionality as familiarity (i.e., being repeatedly experienced) in Holyoak & Stamenković's (2018) survey of major empirically investigated theories of metaphor.

[3]The *target* subject matter is the one that is being addressed, and the *source* subject matter is the one being used to address it. For instance, for '*firm belief*', the source subject matter can be taken to be that of physical objects and their properties, and the target subject matter can be taken to be mental states. (Neither subject

be open to the possibility that something more than simple retrieval happens. Holyoak & Stamenković (2018) postulate that some conventional metaphors may be comprehended by a constrained, non-cognitively weighty analogy mechanism.[4]

2.2.2. *Novelty in general*

We take the common approach of defining novelty in metaphor as a lack of conventionality – the fewer widely shared meanings that a metaphorical expression has, the more novel it is. The more novel it is, the less likely it is that a hearer can just retrieve a familiar, relevant metaphorical meaning, and so the more likely it is that he/she must work out a metaphorical meaning in some other way, a 'beyond retrieval' way as we will say. To the extent that the comprehension of novel metaphor involves a beyond-retrieval process, we are faced with a range of different processes that have been proposed by different metaphor theories. We are also faced with the possibility that different types of processing, whether drawn from the same theory or from different theories, are useful in different types of metaphorical phrases, different contexts and so forth. In defining different varieties of metaphorical novelty, we do not wish to be bound by the limits imposed by particular theories of metaphor. We therefore abstract two general forms of beyond-retrieval processing that are discernible in different specific forms in different theories, and that can therefore plausibly be assumed to be used by hearers in some way, irrespective of what one's specific theory of metaphor comprehension is. These general forms are: (1) the use of *bridges* between source and target subject matters; and (2) the following of *within-source connections*. These are discussed separately in the following two subsections.

2.2.3. *Metaphoric bridges: mappings, superordinate categories etc.*

The comprehension of metaphor must involve some way in which the hearer can take aspects of the source subject matter to relate suitably to aspects of the target subject matter. Different metaphor theories have different versions of this, but we generalise by using the term '[metaphoric] bridge' for whatever device is being used to relate source to target aspects. Some prominent theories are in some way reliant on bridges in the form of 'mappings'. Consider first a metaphor theory that proposes that the hearer finds an analogy entirely from scratch between the source and target subject matters as the central part of comprehension. A salient example here is the structure-mapping theory (Gentner, 1988; Gentner & Wolff, 1997; Wolff & Gentner, 2011); also incorporated as one part of the career of metaphor theory proposed by Bowdle & Gentner (2005). The analogy consists of postulated mappings between aspects of the source and aspects of the target. For instance, if an academic department is talked of metaphorically as a solar system, the hearer might be theorised to create a mapping between the most prominent researcher and the star of the solar system, mappings between other academic staff members and

---

matter has clear boundaries.) We also say the subject matters are on the *source side* and *target side*, respectively.

[4]See also Casasanto & Gijssels (2015) and Hampe (2017) for commentary on some of the evidence amassed by various authors about the source subject matter of metaphors being activated, and Barnden (2020) for further commentary and a discussion of what such activation means for metaphor theory in general.

planets, a mapping between the relationship of working under someone and gravitational attraction to a heavier object, a mapping between academic prowess and physical size of solar-system objects and so forth. In such a theory, the 'bridges' are the mappings created.

Another possibility is that the hearer already knows of an analogy, and extends it for the purposes of comprehending a specific utterance. For instance, we can regard a conceptual metaphor (Lakoff & Johnson, 1980, 1999), such as LOVE RELATIONSHIP AS A JOURNEY as a body of already-known analogy (Holyoak & Stamenković, 2018). That is, it is an already-known set of mappings between love items and journey items (e.g., between the lovers and the journey companions, and between the progress of the relationship and the progress of the journey). However, the comprehension of a particular sentence that relies on such a conceptual metaphor, for example, 'John and Mary were on a lazy tropical cruise together' when this is taken to be metaphorically about John and Mary's love relationship rather than literally about a cruise, might involve the creation of new mappings. An example would be a mapping that puts physical effort (a lack of which is suggested by 'lazy') in correspondence with emotional and mental effort aimed at sustaining the relationship. Similar points apply, though with great differences of specific detail, if the already known mappings are of much more generic sorts such as in the primary metaphors of Grady (1997) (e.g., ORGANISATION IS PHYSICAL STRUCTURE) or in the correspondences proposed by the ATT-Meta theory (Barnden, 2015, 2016; Barnden & Lee, 2001). In the theories of the sort discussed in this paragraph, we take both the already-known mappings and the newly constructed ones as bridges.

Other theories propose other ways whereby hearers relate source aspects to target aspects. Most saliently, consider categorisation-based theories. We include here not just Glucksberg's (2001) theory, but also proposals from Relevance Theory, such as in Carston & Wearing (2011) and Sperber & Wilson (2008). Such theories propose the finding and/or construction of categories (often called 'super-ordinate' categories) that include a mentioned source category and also the target item. Take, for instance, the widely discussed (Glucksberg, 1998) example 'My job is a jail'. Here a category jail* of situations that are behaviour-limiting and unpleasant may be retrieved or invented. This category contains the (real-)jail category as a subcategory, but is also broad enough to include at least some jobs or types of jobs. We count this superordinate category as the bridge used between source and target. As we have already indicated, such a superordinate category may be either already known or freshly invented. An intermediate possibility is that the hearer knows of a superordinate category jail* that includes real jails, but finds it to be too narrow or too broad to fit the specific job with optimal aptness, and so constructs a new category jail** that is broader or narrower, as needed. This makes the point that metaphor comprehension may involve the refinement of an existing bridge as opposed to the construction of an entirely new one. Alternatively, instead of theorising in terms of superordinate categories, one can do so in terms of shared properties. In the job/jail example, the shared properties would be behaviour limitation and unpleasantness. These properties would be aspects of both target and source, and would serve as a type of bridge.

The types of bridges so far mentioned are drawn from the main types of theories surveyed by Holyoak & Stamenković (2018), namely detailed fresh-analogy-finding accounts, conceptual metaphor theories (here considered as a type of known-ana-logy-cum-fresh-analogy-finding theory) and categorisation theories. Their survey

covers those theories that are about adult metaphorical comprehension and that have been subjected to major psychological investigation aimed at support or rebuttal. Their survey also mentions some further possibilities, for example, a weaker, simpler form of analogising. Nevertheless, the sense in which source and target aspects are related is similar to what was already described above. It should be noted that their survey also omits some major theories, notably blending theory (or conceptual integration theory; Fauconnier & Turner, 1998, 2008), presumably on the grounds that it has mainly been subject to linguistic theorising rather than psychological experimentation. The melding of different conceptual items within blend spaces can be seen as a type of bridging that is different from the types above, but in fact such melding is tied to the existence of (already known or newly stipulated) mappings between subject matters as well.

An aspect of our relative neutrality with respect to different theories is the way we hypothesise the existence of particular familiar bridges. Although we have been influenced by specific bridges that existing theories have postulated as familiar, we do not rely only on these postulations. Instead, as we will see below, in our study, we use conventional expressions (including those in our materials as specified in Section 3, but also others) directly as evidence for the existence of bridges. For instance, we used the conventional metaphoricity of *'firm belief'* to evidence a familiar bridge between beliefs and physical objects (the Supplementary Document contains more examples of this).

In summary, bridges are the constructs proposed by theories as constituting the relationships that the hearer sees between source and target subject matters and that help to comprehend the metaphorical utterance at hand. The notion of bridge is not intended as a new idea, but just as a convenient way of abstracting what is common from well-known proposals by previous researchers.

2.2.4. Within-source inferencing and other connection-following

Beyond-retrieval metaphor comprehension can also benefit from something orthogonal to the use of bridges, namely the following of inferential or other connections *within* the source subject matter, whereas bridges are *between* the target subject matter and the source subject matter. As an example, let us assume that, when something is being metaphorically viewed as a source of physical light, there is a familiar bridge between (on the source side) the physical brightness of the item and (on the target side) the item's usefulness and strikingness. Thus, one way to comprehend *'high-wattage idea'* is to assume that the idea is being viewed as an illumination source and then to infer from its high wattage that it is especially bright. This especially strong brightness then plausibly suggests, via the bridge, that the idea is especially useful and striking. However, the inference to the especially strong brightness from the high wattage is entirely within the source subject matter.

Connections followed within the source subject matter do not need to be a matter of inference. For instance, they might be a matter of negation, opposition or lack. Suppose an idea that has so far been considered to be 'bright' is said to be 'extinguished'. This might be comprehended by taking the idea to be an initially bright light source as above that at some point has its shining stopped. The resulting non-shining could be interpreted as non-usefulness, using the same bridge as above. There is certainly an inference here from extinguishing to non-shining, but to use this to

exploit the bridge between shining and usefulness, there needs to be a step taken from non-shining to shining.

The use of inferencing or other connection-following within the source subject matter, or 'on the source side' as we will also say, is discernible in many theories of metaphor, but is especially notable in relevance theory proposals (Carston & Wearing, 2011), ATT-Meta (cited above) and the approach taken by Ruiz de Mendoza and colleagues (Ruiz de Mendoza & Galera, 2014, especially pp. 108ff). Within-source connection-following (inferencing etc.) is also involved in the use of so-called 'entailments' of conceptual metaphors (Lakoff & Johnson, 1980, 1999; Lakoff & Turner, 2009). It therefore spans a number of different theories of metaphor comprehension.

2.2.5. Two types of metaphorical novelty: high-conforming and low-conforming

The particular constellations of beyond-retrieval processing types (based on existing bridges, refinement of existing bridges, construction of new bridges and within-source connections) that are proposed by different theories of metaphor comprehension could have different effects on speed of processing and on the results of meaning construction. Furthermore, for a given expression, markedly different processing mixes might be used in different contexts or by different hearers. In essence, there is an extensive, intricate landscape of different beyond-retrieval processing scenarios that could be proposed to account for the novelty of metaphorical expressions and the ways in which they are processed, and derivatively we can talk of there being different types of novelty on the basis of such scenarios.[5]

The points discussed here are somewhat expanded upon in the Supplementary Discussion. They require much more extensive development than is possible in this paper, but as they stand, they suggest that we can usefully pick out a relatively easily comprehended, *high-conforming* type of novel metaphor. We deem a nonconventional metaphorical expression to be high-conforming if, when the expression is presented out of context, the following holds:

- major *distinctive* aspects of source-side concepts raised by the expression can be used,
- *without creatively entertaining any special context*,
- to provide *distinctive* aspects of a metaphorical meaning for the expression,
- by means only of bridges that can plausibly be suggested to be *familiar to hearers* together with
- *easy, generally applicable* within-source connections.

In dubbing an expression as high-conforming, we assume that different metaphor theories would agree that the expression can be comprehended in the way indicated, even if the particular bridges and within-course connections deployed might be importantly different, and even if the distribution of effort between the use of bridges

---

[5]A slight complication in our discussion is that beyond-retrieval processing could in principle be used for transparent conventional metaphor as well as for novel metaphor. Empirical results such as those summarised in Holyoak & Stamenković (2018) suggest that this does not *generally* happen, and our main interest in the working-out is in application to novel metaphor, but as a matter of principle and of thoroughness, we should not restrict the possibility of beyond-retrieval processing solely to novel metaphor.

and the use of within-source connections is different. Given the comments above, the expression *'high-wattage idea'* is arguably high-conforming as a novel metaphorical expression. In the process imagined above, all that was needed was a simple, generally applicable within-source inference from the distinctive source-side feature of high wattage to especially high brightness, together with the use of a familiar bridge between degree of brightness and degree of strikingness and usefulness; and no special context needed to be creatively entertained.

By contrast, a novel metaphorical expression is *low-conforming* if, when the expression is presented out of context, the following holds:

- in order to derive a *distinctive* metaphorical meaning for the expression that
- exploits *distinctive* source-side concepts raised by the expression,
- the hearer needs to: creatively entertain *special contexts*, AND/OR
- construct *new* bridges (or refine old ones[6]), AND/OR
- follow *difficult or not generally applicable* within-source connections.

We place *'curved hope'* in this category, because of the probable need to entertain a special context, and because of the relative difficulty of making useful within-source inferences from curvedness. Now, a hearer might infer from the curvedness that the hope is probably being viewed as a solid physical object, and then infer that it has at least a normal level of robustness. However, this meaning, relying only on familiar bridges and easy within-source inferencing, does not exploit anything very distinctive about being curved as opposed to being straight. One possible meaning would be to take being curved as opposed to straight as implying greater visual appeal – especially if, let us imagine, the attractiveness of hopes is under discussion in a rather unusual context – and to use some new or old bridge between visual appeal and more abstract attractiveness.[7]

Note that the low-conforming type of novel metaphor is not simply the negation of the high-conforming type – it involves a fairly strong departure from high conformity. Novel metaphors can conform to any degree, and can lie between the two types. These types are intended to lie relatively near either end of the scale of conformity to what is familiar, generic and straightforward. High-conforming novel metaphor covers at least some of what is commonly called extended metaphor, although the more difficult cases of extended metaphor would be low-conforming.

In our characterisations of high conformity and low conformity, we have sought to build in matters other than the question of whether new bridges are needed, in order to respect the point that hearers of metaphorical expressions may derive meanings of different richness depending on how strongly they exploit distinctive source-side features, what special contexts they entertain and what relatively difficult or special within-source inferencing or other connection-following they do.

We should point out that what counts as high conforming or low conforming depends on what counts as a bridge and what counts as within-source inferencing or other connection-following. A metaphor theory that proposes different types of bridges of connection might identify different boundaries between high-conforming

---

[6]Henceforth, for brevity, we will count such refinement as a type of construction of new bridges.

[7]The Supplementary Materials contain more examples of how we decided whether a novel pair was high-conforming or low-conforming.

and low-conforming novel metaphors. Additional types of existing knowledge, beyond knowledge of possible contexts, bridges and within-source connection-following possibilities would also play a role in making this distinction. The study below is confined to our current notions of high conforming and low conforming. However, given that it takes into account the main empirical approaches, as surveyed in Holyoak & Stamenković (2018), it should be reasonably robust in the context of current developments in metaphor theory.

We also need to stress that the high-conforming/low-conforming distinction may well be language-specific in that speakers who know different languages may to some extent deploy different bridges and know different within-source connections. The classifications of expressions in our study are from the point of view of English.

Finally, we do not assume that the process of constructing or using bridges, following inferential or other connections on the source side, or considering possible contexts, needs to be conscious. Nor do we assume that a meaning that is derived for the expression is clearly apparent in consciousness. These points greatly affected our methodology below.

2.3. Hypotheses

We divide the hypotheses for our experiment into two sets. The first set is centred on the differences between monolinguals and multilinguals in making sense of novel metaphor, irrespective of novelty type (high conforming versus low conforming), whereas the second set is centred on the way in which those two different types of novelty affect readers' attempts to make sense of an expression, irrespective of the readers' mono/multilingualism. We had no specific expectation as to whether there would be an interaction whereby multilinguals and monolinguals act more differently from each other on one type of novelty than on the other.

Our main hypotheses concern the likelihood of someone taking a novel metaphorical expression to make sense (be meaningful), as one aspect of the person's degree of cognitive flexibility. As a subsidiary matter, we also include hypotheses about the time taken to make those meaningfulness judgements, as a way of illuminating the amount of effort involved, potentially giving more detailed insight into cognitive flexibility.

We include both literal and conventional metaphorical expressions in our study as something to contrast novel metaphorical expressions to, but the study does not aim to explore differential effects from being literal as opposed to conventional metaphorical. Different theories may differ on whether an expression is literal or conventional metaphorical, and as indicated above, it is reasonable to claim that conventional metaphorical meanings are typically found by simple retrieval from memory, on a par with literal meanings of words. We thus assume that people make sense of conventional metaphorical expressions about as easily as literal expressions, in line with findings from previous research (Werkmann Horvat, Bolognesi, & Lahiri, 2021) However, there is no guarantee that a meaning that could reasonably be claimed to be conventional metaphorical is obtained simply by memory retrieval (see Section 2.2.1), so we still mention literal and conventional metaphorical phrases separately in our hypotheses.

2.3.1. Hypotheses concerning monolinguals vs. multilinguals

Hypothesis A. Multilinguals are more likely to take novel metaphorical expressions to make sense than monolinguals are, regardless of whether the expressions are high-conforming or low-conforming.

Hypothesis B. Multilinguals take longer than monolinguals to assess whether novel metaphorical expressions make sense, whether they are high-conforming or low-conforming.

These hypotheses follow from the discussion in Section 2.1. Since the multilinguals as well as the monolinguals in our experiment are L1 English speakers, the differences could be subtle. In particular, as regards Hypothesis B, the factors affecting timing might be too mixed in their individual effects to show a generalisable overall effect. The extra information potentially possessed by multilinguals could tend to slow them down, while possibly speeding them up in cases where it allows considerably easier comprehension of an expression.

2.3.2. Hypotheses concerning effect of novel metaphoricity

Hypothesis C. Monolinguals and multilinguals are (a) more likely to take expressions to make sense when they are literal or conventional metaphorical than when they are novel metaphorical (whether in a high-conforming or low-conforming way); and (b) more likely to take novel metaphorical expressions to make sense when they are high-conforming than when they are low-conforming.

Hypothesis D. Monolinguals and multilinguals (a) take less time in assessing whether literal or conventional metaphorical expressions make sense than whether novel metaphorical ones (of either type) do so; and (b) take different amounts of time in assessing whether low-conforming as opposed to high-conforming novel expressions make sense.

Hypotheses C(a) and D(a) are based on the suspicion that conventional metaphorical expressions are typically comprehended by simple retrieval of metaphorical meanings, and on findings from experiments suggesting that comprehension of novel metaphors is relatively taxing (Gentner & Wolff, 1997; Lai & Curran, 2013; Lai, Curran, & Menn, 2009; Rutter et al., 2012; Werkmann Horvat, Bolognesi, & Kohl, 2021).

Hypothesis C(b) is motivated by the special work needed for rich comprehension in low-conforming cases. Nevertheless, the hypothesis may be invalidated if enough participants settle for less-rich comprehension of some low-conforming expressions, or enough participants fail to comprehend some high-conforming expressions.

As regards hypothesis D(b), there are considerations that pull in different directions as regards processing speed. Low-conforming expressions are more likely than high-conforming ones to need, for rich comprehension, the creative entertainment of special contexts, the construction of new metaphoric bridges (e.g., new/refined mappings or superordinate categories) or the finding of unusual within-source inferential or other connections. Therefore, attempts at rich comprehension of a low-conforming expression, exploiting distinctive aspects of the source-side concepts raised, may often be slower than for a high-conforming one (recall our discussion of *'curved hope'* in Section 2.2.5). However, precisely because of the relative difficulty of rich comprehension in low-conforming cases, participants may, more frequently than in high-conforming cases, give up trying to comprehend at all or settle for less-rich, quick-to-derive comprehension that ignores distinctive aspects of source-side

concepts and eschews special contexts or new metaphoric bridges. A faster response than in high-conforming cases could even arise. Nevertheless, due to the assumed major differences in comprehension, we predict that there will be an overall speed difference. It was our aim to explore the direction of this difference in this study, as a basis for further studies.

3. Methods

The experiment, in which sentences containing literal and metaphorical expressions were presented to participants, was a self-paced reading study combined with a YES/NO meaningfulness judgement. We recorded answers to the meaningfulness judgement, the time it took to make the judgement and the time needed to read each region of the sentence.

   This type of task was chosen because we were interested in seeing possible comprehension difficulties upon encountering a certain phrase, which this task allowed partly through measuring the reading time in certain sentence regions, as well as observing the polarity of the answer to the meaningfulness question and measuring the time taken to make the answer. We assume that reading time here acts as a proxy for processing difficulty, and therefore we assume that larger reading times can be interpreted as reflecting difficulty in finding meaning, since other effects that might cause processing difficulties are controlled for. Thus, we considered the notion of 'making sense' of an expression to involve not just the making of the meaningfulness judgement, but also processes undertaken while reading the sentence.

   Notably, we did not ask participants to report meanings that they discerned for the sentences. As discussed further in Section 5.3, the main reason for this was that we did not wish to assume that participants were necessarily conscious of the meanings, or at any rate to have a clear and full enough conscious awareness of them to allow useful reporting of them under the pressure of an experiment (even though we imposed no time limit on answering the meaningfulness question). This generates some problems in interpreting results (see Section 5.3), but, on the other hand, it properly allows for the fact that some of our novel examples, especially but not exclusively some low-conforming ones, are difficult to assign a specific meaning to; and even when one does have a specific meaning consciously in mind, it can be difficult to express. Furthermore, given the relatively high indeterminacy of meaning of novel metaphor, a participant may detect a range of different meanings, and therefore face a difficult task if required to choose between them or to summarise them in the service of reporting a meaning. These points also mean that a meaningfulness 'judgment' by a participant might not be anything like a considered judgment (result of deliberation) but might simply be based on a feeling of comprehending or failing to comprehend.

   All data, stimuli, analyses and supplementary materials are stored in an online repository on the Open Science Framework (OSF): https://osf.io/ek4q8/?view_only= faa82d8334fd478cb8f99fd15f107597.

   The research was reviewed by, and received ethics approval through, the University of Oxford Central University Research Ethics Committee (Ethics Approval Reference: R56945/RE001).

3.1. Stimuli

3.1.1. General characteristics

The metaphorical expressions were English adjective–noun pairs (A–N pairs; henceforth, referred to simply as *pairs*), as in *'firm belief'* and *'hazy hope'*. We assumed that the most likely metaphorical meanings would arise from the noun being taken literally, with only the combination with the adjective making the expression metaphorical. For instance, we assumed a *'shaky price'* would be taken as meaning a financial price that is metaphorically shaky, not as a metaphorical price (e.g., diminution of health) that is (metaphorically) shaky. Hence, in the application of the notions of high conforming and low conforming to the novel pairs, the noun and adjective were assumed to indicate the target and source subject matters, respectively.

The critical stimuli consisted of 96 pairs (see Table 1), based on 24 different nouns (four for each of six target subject matters: mind, time, economics, weather, geography and food-and-drink). For each noun, there were four different pairs, for four different metaphoricity conditions: a literal pair and three metaphorical pairs – a conventional one (in an operationalised sense explained below), a high-conforming novel one and a low-conforming novel one. No adjective was used with more than one noun. For instance, the four pairs using *idea* were *'simple idea'* (literal)*, 'bright idea'* (conventional metaphorical)*, 'grey idea'* (high-conforming novel metaphorical) and *'damp idea'* (low-conforming novel metaphorical), and the four adjectives here appeared with no other noun.

Table 1. *Critical nouns and adjectives*

Nouns	Adjectives			
	Literal	Conventional	High-conf	Low-conf
Idea	Simple	Bright	Grey	Damp
Pride	Ethnic	Foolish	Wise	Edible
Hope	False	Faint	Hazy	Curved
Belief	Sincere	Firm	Thick	Fitted
Year	Rainy	Golden	Brass	Locked
Night	Chilly	Lengthy	Fat	Pointed
Week	Current	Quiet	Muted	Liquid
Hour	Crucial	Peak	Bottom	Milky
Loan	Useful	Flexible	Stiff	Purple
Fee	Rental	Fixed	Loose	Pale
Tax	Local	Heavy	Slim	Wet
Price	Fair	Stable	Shaky	Melted
Storm	Wintry	Raging	Kicking	Blond
Cloud	Ugly	Angry	Furious	Loyal
Rain	Freezing	Gentle	Excited	Witty
Wind	Warm	Fierce	Rude	Married
Creek	Tiny	Roaring	Howling	Drunk
Hill	Steep	Lonely	Troubled	Tied
River	Shallow	Lazy	Tired	Dizzy
Lake	Salty	Calm	Eager	Ripe
Beer	Fizzy	Strong	Tough	Folded
Tea	Fragrant	Weak	Fragile	Keen
Wine	Tasty	Bold	Humble	Silent
Soup	Instant	Hearty	Shy	Shouting

Each of our 96 pairs was embedded in a short sentence. These critical sentences were read by participants in the self-paced reading task. The sentences were partitioned into four nonoverlapping lists, each of which contained 24 of the 96 critical sentences, plus 72 filler sentences containing only nouns and adjectives not used in the critical sentences. Each list contained all 24 critical nouns, hence each noun only once. A given noun appeared in different metaphoricity conditions in different lists. A given participant saw just one of the four lists. The metaphoricity conditions were evenly spread across each list (six critical sentences for each of the four conditions), and the 48 participants were evenly spread across the lists.

For each of the four pairs involving a given noun, the surrounding wording in the sentences containing the pairs was the same. See the examples in Table 2.

The reason for embedding the pairs in sentences was to provide reasonably natural syntactic contexts and to avoid possible special processing effects at the start and end of reading a word sequence. We acknowledge that including the adverb at the end of the sentence might affect participants' judgements about the A/N pair. However, the adverbs were always the same across the four conditions for any given noun, which somewhat controls for this effect. The final adverb was added to track possible spillover effects from reading the noun.

Note that the definitions of high-conforming and low-conforming novelty in Section 2.2.5 are predicated on the expression being presented out of context. The presentation of our pairs therefore departs slightly from such pure presentation. However, we endeavoured to keep the wording outside the pairs in the sentences semantically generic so that it would give participants relatively little help towards finding particular meanings for the pairs, under the above assumption that the participants would take the nouns literally. For instance, in Table 2, the 'idea' is 'suggested quickly' and the 'loan' is 'got easily': whatever the nature of the idea, it might be suggested quickly under suitable circumstances, and whatever the nature of the loan, it might be got easily under suitable circumstances. The wording may direct participants towards certain interpretations, but we felt the sentences were a suitable compromise between the ideal of reasonably natural syntactic contexts and the ideal of null semantic contexts.

As shown in Table 2, each sentence was divided into six regions. The presentation of each sentence was followed by the meaningfulness question, 'Does this sentence make sense?' with a forced choice between YES and NO.

The adjectives were balanced for frequency and length across the four conditions. A single-factor analysis of variance showed that there were no significant differences for frequency and length. See Table 3 for details.

Table 2. Examples of sentences

Condition	Region 1	Region 2	Region 3	Region 4 (Adj)	Region 5 (Noun)	Region 6 (Adv)
Literal	John	suggested	a	simple	idea	quickly
Conventional	John	suggested	a	bright	idea	quickly
High-conf	John	suggested	a	grey	idea	quickly
Low-conf	John	suggested	a	damp	idea	quickly
Literal	Cath	got	a	useful	loan	easily
Conventional	Cath	got	a	flexible	loan	easily
High-conf	Cath	got	a	stiff	loan	easily
Low-conf	Cath	got	a	purple	loan	easily

Table 3. Lexical measures for adjectives

	CELEX freq	Length
Mean lit	44.69 (*SD* = 59.19)	5.75 (*SD* = 1.22)
Mean conv	43.01 (*SD* = 40.66)	5.54 (*SD* = 1.18)
Mean high-conf	29.78 (*SD* = 25.76)	5.29 (*SD* = 1.36)
Mean low-conf	31.02 (*SD* = 28.25)	5.29 (*SD* = 1.20)
F	(3, 92) = 0.88	(3, 92) = 0.76
p	0.452	0.519

3.1.2. Literal and conventional metaphorical pairs

We checked the following when judging a pair to be literal: that our chosen dictionaries,[8] taken collectively, gave a sense for the adjective and a sense for the noun such that: these senses were current, and not dependent on specialised knowledge; neither sense was listed as metaphorical, figurative, dialect, slang and so forth; the adjective's sense applied directly and easily to the noun's sense and the resulting composed meaning for the pair was directly usable in the sentence the pair was embedded in.

Our conventional pairs were not required to be conventional metaphorical terms in themselves, although some were. Rather, we dubbed a pair as a conventional metaphorical one for the purposes of our study when the *adjective* by itself had a current, nonspecialised conventional metaphorical sense that applied directly and easily to a current, nonspecialised literal sense for the noun, and delivered a composed meaning for the pair that was directly usable in the embedding sentence. We also required a minimum corpus frequency of 10 for operational reasons. We judged an adjective to have a particular conventional metaphorical meaning when that meaning was listed in one or more of our dictionaries and we judged it to be metaphorical. In a few cases, our judgment was aided by a sense being marked by the dictionary as metaphorical or figurative. In either case, we required the meaning to be transparent (in our judgment) in the sense of Section 2.2.1.

We included conventional pairs in our study because we were interested in seeing whether at least some conventional pairs, even though transparent, were easier and quicker to comprehend by our participants than the high-conforming novel pairs. If so, we would have evidence that the conventional pairs were being comprehended through simpler processing so that novel metaphoricity was actually having an effect.

It was partly to maximise the possibility of obtaining this evidence that we also included literal pairs and ensured that our conventional pairs were, overall, roughly as familiar as our literal pairs. We therefore counted corpus occurrences of the pairs in the 45-billion-token English Web Corpus (enTenTen) using SketchEngine. We found that our literal and conventional pairs occurred in similar numbers overall. A paired *t*-test showed that the difference between the literal and conventional conditions was nonsignificant (*M*Lit = 2,844.5, *SD* = 4,290.45; *M*Conv = 1,898.08, *SD* = 2,778.01; *t* = 0.94; *p* = 0.355). Note that a literal or conventional pair in our study could be *lexically* novel in the sense that it did not occurr frequently. In particular, for a literal pair to be quickly and easily

---

[8]The Chambers Dictionary, 2003, the Oxford English Dictionary Online, full version, accessed in 2020 and 2021, and Webster's Third New International Dictionary, unabridged, 1961.

comprehended by a hearer it was not necessary for it to ever have been encountered before by that hearer; a similar latitude also needed to apply to the conventional pairs.[9] Therefore, the contrast in metaphorical novelty between (1) the literal and conventional pairs, on the one hand, and (2) the novel pairs, on the other hand, was not on the basis of *lexical* novelty, but on the basis that the pairs in did not contain *metaphorical* novelty: the literal pairs were intended to be nonmetaphorical, and the metaphoricity of a conventional pair was intended to be nonnovel metaphoricity arising from its adjective. Having said this, the literal and conventional pairs were in fact, overall, much more frequent in the corpus than the novel pairs, as will be explained in Section 3.1.3.

3.1.3. Novel metaphorical pairs and occurrence frequencies

A basic requirement for the novel pairs, as used in our sentences, was that they were not classed as literal or conventional metaphorical as above. We made a judgment about whether a novel pair (that was metaphorical in our judgment) was low-conforming or high-conforming on the basis of: (1) evidence afforded by conventional metaphorical expressions about the existence of familiar bridges (see Section 2.2.3); and (2) our judgments about what special contexts, if any, needed to be considered and what within-source inferencing or other connection tracing was needed. Some detailed examples of our decision making are given in the Supplementary Document.

As regards corpus frequency, we considered that metaphorical novelty would be most likely to be present if the pair was lexically rare. We therefore severely limited the number of allowed occurrences in the English Web Corpus (as determined by SketchEngine). In the low-conforming case, we did not allow there to be any occurrences in the corpus at all. In the high-conforming case, we considered that it was acceptable for there to be a few occurrences, because of the familiarity of bridges (e.g., mappings) used, the straightforwardness of within-source inferencing or other connection-following, and there being no need for considering unusual contexts. However, for safety, we still required occurrences to number below 10. Overall, the mean level of occurrence was $M = 1.58$ ($SD = 1.98$). Note, however, that low-conforming and high-conforming novel pairs do not need to have these degrees of rarity in general.

Although our nonnovel pairs were not required to be frequently occurring, their frequencies in the English Web Corpus (accessed via SketchEngine) were in fact much higher overall than those of the novel pairs. The mean frequencies as above were: 2,844.5 for literal, 1,898.08 for conventional, 1.58 for high-conforming novel as noted above and of course zero for low-conforming novel.

Since the standard deviations in the literal and conventional cases were also large, the ranges of frequency need to be considered. The important comparison here is between the highest frequencies for high-conforming pairs and the lowest frequencies for literal and conventional ones. The high-conforming pairs had a maximum frequency of only eight occurrences, reached by one pair (*'rude wind'*), with all the remaining pairs at four occurrences or below. In fact, 10 of the 24 high-conforming

---

[9]Of course, normally, the adjective will frequently have been used with its conventional metaphorical meaning in other phrases before. Recall also that we required at least 10 corpus occurrences, although this did not enforce a marked difference to the literal phrases as the minimum frequency of these turned out to be 6.

pairs had no occurrences at all. By contrast, although we allowed conventional pairs to have a frequency as low as 10 (this number chosen operationally to ensure a break with high-conforming pairs; see the Supplementary Document for more information on this), the 24 conventional pairs in fact had a minimum frequency of 21 (*'lengthy night'*), with only 2 other pairs having 100 occurrences or fewer (*'bold wine'* with 44 and *'angry cloud'* with 97 occurrences). The 24 literal pairs had a minimum frequency of 6 (*'useful loan'*), which is comparable to the maximum of 8 for high-conforming pairs, but only 3 other literal pairs had 100 occurrences or fewer (24, 65 and 88 occurrences).

Although it has been fairly common in previous metaphor studies to assess the degree of conventionality or novelty of metaphorical pairs by acquiring ratings from participants (separate from those taking part in the main experiment), in this work, we moved away from this method. The main reason for this is that we could not expect nonexperts to assess whether a metaphorical pair was novel in a high-conforming way or in a low-conforming way. With respect to the distinction between metaphorically novel and nonnovel pairs, we suggest that is more reliable to rely on (non)novelty as revealed by dictionary definitions than on participants' within-experiment and conscious opinions about the matter. This is because such opinions might be unduly influenced by lexical novelty, whereas our notion of novelty only indirectly involves lexical novelty.

3.1.4. Norming study concerning comprehension difficulty

We naturally expected L1 English speakers in general to find low-conforming novel A–N pairs to be more difficult to comprehend than high-conforming ones, and the latter to be more difficult than nonnovel (literal or conventional) ones.

Accordingly, as a further, indirect, check on our classification of our metaphorical pairs, we normed the pairs with L1 English speakers using Qualtrics, after all materials had been created and classified. We included literal pairs for completeness. We recruited 130 L1 English speakers (different from any in the main part of the experiment), who each saw all 96 pairs. They were asked to assess the comprehensibility of each pair, on a Likert scale. The instructions were:

'For each phrase mark how easy is it to understand the phrase on a scale from 1 to 7:

1 being the most difficult, 7 being the easiest'.[10]

As Table 4 shows, there was a highly significant difference in participants' ratings of literal and conventional *versus* both high-conforming and low-conforming novel pairs, and in the ratings of high-conforming *versus* low-conforming pairs. The difference between literal and conventional was also significant although small (the literal pairs were slightly more comprehensible than the conventional ones).

---

[10]We assumed that for ordinary English speakers, the word 'understand' would be more accessible and natural than the word 'comprehend' used in the text of the present paper. We take these words meanings to be closely similar.

Table 4. Results of the norming study (paired T-test)

Conditions	Mean	$P(T <= t)$ two-tail	t-Stat	Cohen's d
Lit vs. conv	lit = 6.62 (SD = 0.30) conv = 6.32 (SD = 0.54)	0.021	2.48	0.68
Lit vs. high-conf	lit = 6.62 (SD = 0.30) hi-c = 3.20 (SD = 0.92)	<0.0001	−17.23	4.99
Lit vs. low-conf	lit = 6.62 (SD = 0.30) lo-c = 1.89 (SD = 0.35)	<0.0001	57.68	14.51
Conv vs. high-conf	conv = 6.32 (SD = 0.54) hi-c = 3.20 (SD = 0.92)	<0.0001	14.29	4.13
Conv vs. low-conf	conv = 6.32 (SD = 0.54) lo-c = 1.89 (SD = 0.35)	<0.0001	35.79	9.73
High- vs. low-conf	hi-c = 3.20 (SD = 0.92) lo-c = 1.89 (SD = 0.35)	<0.0001	7.34	1.88

3.2. The main experiment

3.2.1. Participants

General characteristics and group classification. All 48 participants were L1 speakers of English, aged 18–55, recruited in Oxford, UK, right-handed, with normal or corrected-to-normal vision and no known language-related, neurological or hearing disorders. They were divided into two groups: those that spoke another language fluently (n = 24), and those who did not (n = 24). The latter group we call monolingual throughout the paper, whereas the former we call *multilingual*, which we take to include bilingual. The same group of participants participated in, and was similarly described in, Werkmann Horvat, Bolognesi, & Kohl (2021)). The number of participants was determined based on the number of the experimental stimuli lists and on previous studies where two groups of participants are compared (Chen & Husband, 2018), but also on the feasibility of recruiting participants with the desired language history.

Forming such groups of participants is a challenging task. One challenge in Europe is finding a true, minimalist monolingual, someone who has only ever been exposed to one language. Moreover, we recruited mainly Oxford University students, who have usually been exposed to at least one non-English language in their primary education. On the other hand, for multilingualism, we did not wish to take the maximalist line of requiring full, L1-like proficiency in one or more non-English languages.

We formed the groups based on self-report. All participants needed to report as being L1 English speakers. The multilinguals and monolinguals were then those who self-reported as, respectively, speaking at least one further language fluently or not doing so. In the advertisement recruiting the participants, we requested participants whose native language is English and who either spoke another language fluently or did not speak another language fluently.[11]

---

[11]We acknowledge that the term *native speaker* is complex and layered (Davies, 2003); nevertheless, we decided that this term sounds most familiar and straightforward for naïve speakers with no linguistic training. Therefore, in the questionnaires, they were asked about their native language(s), rather than their first language.

The term *fluency* was also used as a term that tends to be meaningful for naïve speakers, meaning *the ability to use a given language independently and skilfully in a variety of contexts*. This was the definition given to potential participants if they were not sure about what the term encompasses.

Since our multilingual group included speakers of different languages, it was difficult to test the language knowledge of all participants. Although we acknowledge possible issues with self-reporting such as subjectivity and possible problems with comparability (Tomoschuk, Ferreira, & Gollan, 2019), self-report questionnaires tend to be quick and practical, and also have been shown to correlate with more objective measures, for example, cloze task scores (Sabourin, Brien, & Burkholder, 2014; Tomoschuk, Ferreira, & Gollan, 2019).

According to the self-reporting, some multilinguals were balanced bilinguals, while some spoke more than one non-English language fluently. Thus, our multilinguals were very similar to what Butler (2013) calls multilanguage users, who are either bilingual or fluent in more than two languages. This is appropriate to the current study, which is into how knowledge of at least one non-English language might affect comprehension.

Language questionnaires. To define our two groups, we administered a language questionnaire before the self-paced reading experiment. It was adapted from one used in the ERPLing Lab at the University of Ottawa (Sabourin et al., 2016).[12]

In the monolingual group, the average age was 22 (min 18, max 46, $SD = 6.86$), with 5 males and 19 females. All members identified English as their only native language, and as their parents' native language. Four identified their strongest L2 to be at an intermediate level, with the rest stating it was at a low or very low level. The average age at which they began learning their strongest L2 was 10.5 years, and the average daily use of it was 0.2 hours. The average number of non-English languages they reported knowing was 1.9.

As for the multilinguals, the average age was 28 (min 19, max 52, $SD = 9.52$) with again 5 males and 19 females. All identified English as one native language, whereas 10 identified a further native language. Fourteen reported having a parent whose native language was not English or who was multilingual. All the multilinguals, except one, identified their strongest L2 (or in some cases their second L1) to be at advanced, near-native or native level. One reported two native languages but marked one as being at an intermediate level. Out of all the multilinguals, two participants reported knowing three languages at an advanced level. The average age of acquisition of strongest L2 was 5.5 years, with average daily use at 2.2 hours. On average, the multilinguals reported knowing 3.1 languages beyond English. Non-English languages that the participants reported knowing to an advanced level included: French (6); German (6); Mandarin (2); Portuguese (2); Spanish (2) and Arabic, Bahasa, Cantonese, Fuzhou, Korean and Russian (1 each).

3.2.2. Procedure

The experiment was conducted in a soundproof room in the Language and Brain Laboratory at the University of Oxford. It used the Presentation® software, and the sentences were shown on a Dell Latitude 7480 laptop screen. The participants used a Logitech Gamepad F310 joystick to progress from one sentence to the next, and for each sentence pressed buttons to read through the words at their own pace and then answer the YES/NO meaningfulness question.

---

[12]The questionnaire that was used can be found at: https://osf.io/ek4q8/?view_only=faa82d8334fd478cb8f99fd15f107597.

The self-paced reading experiment lasted about 10 minutes, depending on the participants' speed. The instructions were:

> The words will appear one by one, as you press the button A. This means that you are controlling the reading pace. After every read word, press A to see the next word. Continue with this until the end of the sentence. After that, a question will appear on the screen: Does this sentence make sense? You will have to press either a YES button or a NO button.

4. Results and statistical analysis

4.1. Analysis methods

We analysed the self-paced sentence-reading times, meaningfulness answers and times taken to give the meaningfulness answers of the two participant groups: monolinguals and multilinguals. There were four critical conditions for presented sentences: literal; conventional metaphorical; high-conforming novel metaphorical and low-conforming novel metaphorical.

There were six sentence regions (see Table 2) plus the YES/NO answer region. We analysed how long participants took to read the noun region (Region 5) and the adverb region (Region 6). We also analysed the times taken to answer the question *Does this sentence make sense?* (question/answer region RT, labelled as RTQ below), and the answers themselves (YES/NO). Regions 1–4 were not analysed: Regions 1–3 were the same across conditions (for any given noun), and in Region 4, only the adjective of the adjective–noun pair had appeared, so the pair's metaphoricity was not clear yet.

We excluded responses in the answer region where no answer button was pressed. This resulted in exclusion of 2.6% of data in the answer region. All fillers were excluded from the analysis. The data were analysed using a linear mixed-effects model for RTs and a general linear model with a binomial distribution for meaningfulness answers with the lme4 package, version 1.1-26 (Bates et al., 2015) in R, version 4.0.4. (R Development Core Team, 2011). The fixed effects were condition and group with random effects of participant and item. Initially, the models included by-participant varying intercepts and by-participant varying condition slopes, and by-item varying intercepts and by-item varying group slopes (Winter, 2019). For the noun and answer regions, the model failed to converge with random slopes, and therefore, it was simplified. In the adverb region, an error message suggested the model was overfitted, so the random slopes were also simplified. Contrasts between conditions were analysed using the emmeans R package, version 1.5.4 (Lenth, 2021), and *F*-tests for main effects were analysed using the lmerTest R package, version 3.1-3 (Kunzetsova, Brockhoff, & Christensen, 2017; Lenth, 2021).

4.2. Results

Figs. 1 and 2 display the average reaction times in different regions for the monolingual group (Fig. 1) and for the multilingual group (Fig. 2). The two graphs show partially different trends, which suggest slightly different processing patterns for the two groups.
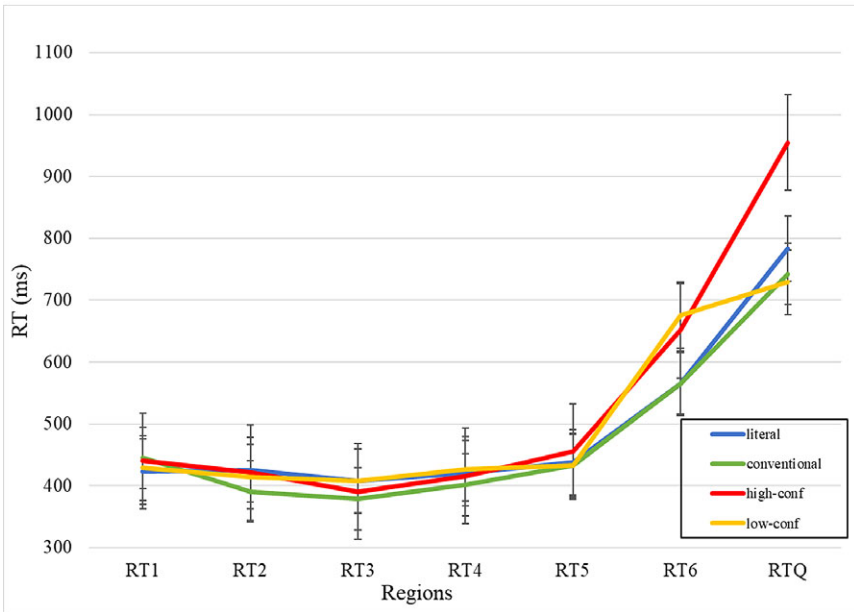
Fig 1. Average RTs in different regions for the monolingual group (SE error bars).
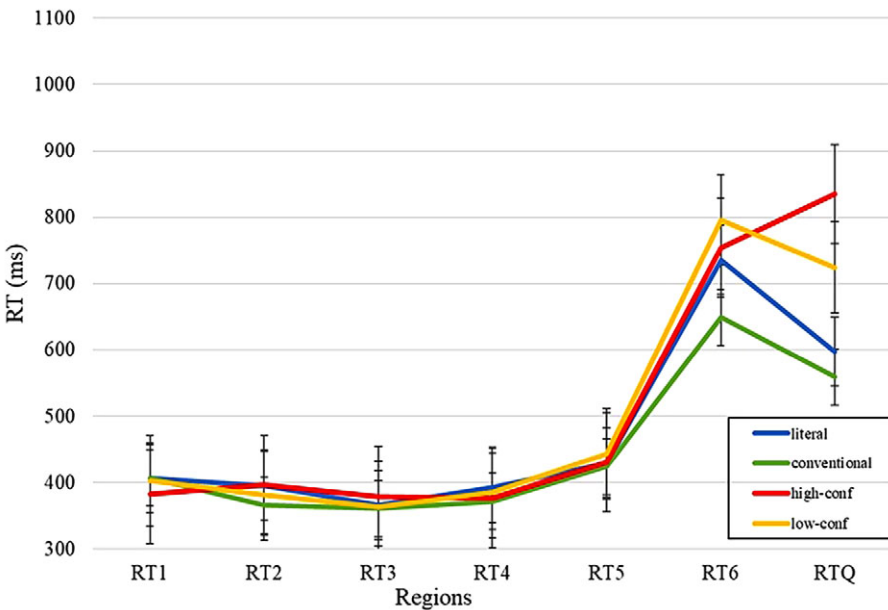

Fig 2. Average RTs in different regions for the multilingual group (SE error bars).

In the noun region (Region 5), there were no significant main effects of condition ($F(3, 1,080.9) = 0.50$, $p = 0.69$) or group ($F(1, 48) = 0.05$, $p = 0.83$). The interaction was also nonsignificant ($F(3, 1,080.9) = 0.65$, $p = 0.58$). In other words, the different stimuli were processed in very similar ways, and the different groups performed in very similar ways.

In the spillover region, that is, the adverb region (Region 6), there was a significant main effect of condition ($F(3, 1,080.8) = 4.71$, $p = 0.003$), but the main effect of group ($F(1, 48) = 1.61$, $p = 0.21$) and the interaction ($F(3, 1080.8) = 0.51$, $p = 0.68$) were nonsignificant. To identify the specific differences between processing times for the stimuli, we ran a post-hoc analysis for the significant main effect of condition. The Tukey's honestly significant difference test with the Kenward–Roger degrees for freedom method showed significant contrasts between high-conforming and conventional ($t(1,087) = -2.58$, $p = 0.049$) and between low-conforming and conventional ($t(1,087) = -3.47$, $p = 0.003$), with high-conforming and low-conforming being slower in each case.[13]

In the answer region, there was again a significant main effect of condition ($F(3, 1,052.68) = 8.70$, $p < 0.001$), but no significant differences for group ($F(1, 48.06) = 1.59$, $p = 0.21$) or interaction ($F(3, 1,052.68) = 1.28$, $p = 0.28$), as in the previous case. In a post-hoc analysis for the significant main effect of condition, the Tukey test as above showed significant contrasts between high-conforming and literal ($t(1,058) = 3.81$, $p = 0.0009$), between high-conforming and conventional ($t(1,058) = -4.83$, $p < 0.0001$) and between high-conforming and low-conforming ($t(1,059) = 3.11$, $p = 0.01$). All models' results are reported in Table 5.

The percentages of YES and NO meaningfulness responses for groups and conditions are shown in Table 6. We ran a generalised linear model with a binomial distribution (Table 7). The model[14] showed a significant main effect of condition ($\chi^2(3, N = 1,118) = 479.38$, $p < 0.0001$) and group ($\chi^2(1, N = 1,117) = 15.86$, $p < 0.0001$), while the overall interaction was nonsignificant ($\chi^2(3, N = 1,114) = 3.06$, $p = 0.38$). *p<0.05; **p<0.01; ***<0.001.

Based on these main effects, we ran simple contrast tests with the Tukey method for p-value adjustment. The results are given on the log odds ratio scale. For condition, the only nonsignificant contrast was conventional-*vs*-literal ($z = -1.45$, $p = 0.47$). The other contrasts were highly significant: literal-*vs*-high-conforming ($z = -9.51$, $p < 0.0001$), literal-*vs*-low-conforming ($z = 14.19$, $p < 0.0001$), conventional-*vs*-high-conforming ($z = 9.13$, $p < 0.0001$), conventional-*vs*-low-conforming ($z = 14.38$, $p < 0.0001$) and low-conforming-*vs*-high-conforming ($z = 7.68$, $p < 0.0001$), with more YES answers in the high-conforming condition. For group, there was a significant monolingual-*vs*-multilingual contrast ($z = -3.03$, $p = 0.002$), with multilinguals more frequently answering YES.

As noted, the condition-group interaction was nonsignificant, and so one would not normally run a post-hoc test contrasting how the different groups performed on the different conditions. In this case, we feel it is justified to look at these contrasts, not with reference to interaction, but to see for which of the four conditions a difference

---

[13]The contrast method for multiple comparisons was *pairwise*. For factor levels A (conventional), B (high-conforming), C (literal) and D (low-conforming), emmeans pairs() function generates the comparisons A–B, A–C, A–D, B–C, B–D and C–D.

[14]For main effects, we used R's anova (glm, test = "Chisq") function because it is recommended in the literature to use a chi-square instead of the *F*-distribution for binomial distributions (Baayen, 2008, p. 218).

Table 5. Linear mixed-effects model results (RTs) for the three regions of interest

Region	Predictor	Estimate	Std. error	df	t value	Pr(>|t|)
Noun	(Intercept)	432.88	28.01	69.35	15.45	<0.001***
	Conditionhigh-c	22.30	17.61	1,080.88	1.27	0.21
	Conditionliteral	5.10	17.61	1,080.88	0.29	0.77
	Conditionlow-c	−0.35	17.61	1,080.88	−0.02	0.98
	Groupmultilingual	−8.98	39.15	66.60	−0.23	0.82
	Conditionhigh-c*Groupmultlingual	−15.25	24.90	1,080.88	−0.61	0.54
	Conditionliteral*Groupmultilingual	0.75	24.90	1,080.88	0.03	0.98
	Conditionlow-c*Groupmultilingual	19.44	24.90	1,080.88	0.78	0.43
Adverb	(Intercept)	565.22	74.35	73.99	7.60	<0.001***
	Conditionhigh-c	86.19	52.33	1,080.84	1.65	0.10
	Conditionliteral	−0.22	52.33	1,080.84	−0.004	0.99
	Conditionlow-c	110.04	52.33	1,080.84	2.10	0.04*
	Groupmultilingual	83.73	104.51	72.63	0.80	0.43
	Conditionhigh-c*Groupmultlingual	18.97	74.00	1,080.84	0.26	0.80
	Conditionlitera*Groupmultilingual	87.03	74.00	1,080.84	1.18	0.24
	Conditionlow-c*Groupmultilingual	37.23	74.00	1,080.84	0.503	0.61
Answer	(Intercept)	724.23	82.76	107.26	8.75	<0.001***
	Conditionhigh-c	243.72	76.12	1,051.82	3.20	0.001**
	Conditionliteral	65.96	76.12	1,051.60	0.87	0.39
	Conditionlow-c	4.90	75.97	1,051.46	0.06	0.95
	Groupmultilingual	−165.75	115.14	103.68	−1.44	0.15
	Conditionhigh-c*Groupmultlingual	35.31	107.88	1,052.30	0.33	0.74
	Conditionlitera*Groupmultilingual	−22.70	107.74	1,051.49	−0.21	0.83
	Conditionlow-c*Groupmultilingual	172.10	108.55	1,053.92	1.58	0.11

*p<0.05; **p<0.01; ***p<0.001.

Table 6. YES or NO meaningfulness answers in percentages for groups and conditions

	Monolingual		Multilingual	
	YES	NO	YES	NO
Literal	92.20 (130)	7.80 (11)	94.29 (132)	5.71 (8)
Conventional	88.81 (127)	11.19 (16)	90.91 (130)	9.09 (13)
High-conf	43.97 (62)	56.03 (79)	59.71 (83)	40.29 (56)
Low-conf	7.80 (11)	92.20 (130)	27.82 (37)	72.18 (96)

Table 7. Generalised linear model for answers to the meaningfulness question

Predictor	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	2.07	0.27	7.81	<0.001***
Conditionhigh-c	−2.31	0.32	−7.35	<0.001***
Conditionliteral	0.40	0.41	0.97	0.33
Conditionlow-c	−4.07	0.37	−10.98	<0.001***
Groupmultilingual	0.23	0.39	0.59	0.56
Conditionhigh-c*Groupmultlingual	0.40	0.46	0.88	0.38
Conditionliteral*Groupmultilingual	0.10	0.62	0.16	0.87
Conditionlow-c*Groupmultilingual	0.81	0.51	1.59	0.11

*p<0.05; **p<0.01; ***<0.001.

between how the monolinguals and multilinguals answers contributed significantly to the main effect of group. The Tukey post-hoc test shows that there are significant monolinguals-vs-multilinguals contrasts in both the high-conforming ($z = −2.62$, $p = 0.009$) and low-conforming ($z = −3.23$, $p = 0.001$) conditions, with multilinguals more frequently answering YES in both conditions (cf. Table 6). In the literal ($z = −0.69$, $p = 0.49$) and conventional ($z = −0.59$, $p = 0.58$) conditions, there is no significant contrast between the groups.

5. Discussion

5.1. Comparison of results to the hypotheses

Overall, the results suggest that being multilingual makes an important difference to the comprehension of novel metaphor of both types (high-conforming and low-conforming), and that the distinction between these two types of novel metaphor is empirically important. In stating this, we should re-emphasise the point made in Section 2.2.5 that the classification in our study of metaphorical expressions as novel or conventional, and, when novel, as high-conforming or low-conforming, was influenced by the nature of English, and that the notions of monolingual and multilingual in our study were relative to English.

Before we consider our specific hypotheses (Section 2.3), we can see that the YES/NO meaningfulness answer profile across the different conditions was highly consistent with the norming study of perceived comprehension easiness (see Table 4). As per Table 6, for both monolinguals and multilinguals, most answers (90% or so) were YES for literal and conventional pairs, fitting the easiness of around 6.5 out of 7 in the norming. Roughly, half (40% for monolinguals, 57% for multilinguals) of the answers were YES for high-conforming, fitting with the medium easiness of 3.2/7 in the norming. Relatively, few answers (around 10% for

monolinguals, 27% for multilinguals) were YES for low-conforming, fitting the low easiness of 1.9/7.

5.1.1. Hypotheses A and B, concerning the effect of multilingualism

Hypotheses A and B concern the effect of multilingualism as opposed to monolingualism on meaningfulness judgements about novel metaphorical expressions (whether high-conforming or low-conforming).

Hypothesis A was supported. The hypothesis is that multilinguals are more likely than monolinguals to take novel metaphor (of either type) to make sense. The outcome is consistent with previous literature that suggests greater cognitive flexibility in multilinguals (see discussion in Section 2.1). Our results show not only that this advantage applies in the particular task of processing novel metaphor, but that it applies both to the processing types used for high-conforming expressions and to the partially different types for low-conforming ones. This therefore suggests that this flexibility in multilinguals is at least present in linguistic tasks, even if our results were taken not to contribute directly to evidence about general cognitive flexibility.

Hypothesis B is that multilinguals take longer than monolinguals to assess whether novel metaphorical expressions make sense, irrespective of whether these expressions are high-conforming or low-conforming. This was not supported. In our results, the time taken to give a YES/NO answer for novel metaphorical expressions did not differ significantly between the two groups. The same lack of significant difference applies to the noun-region or adverb-region reading times. The possible factors that might have played a distinctive role in multilinguals' reaction times were perhaps not strong enough to make a significant difference because both groups were composed of L1 English speakers, as previously suggested in Section 2.3.1.

5.1.2. Hypotheses C and D, concerning the effect of metaphoricity condition

Hypotheses C and D concern the effect of novel metaphoricity on meaningfulness judgements (regardless of whether the judgements are by monolinguals or multilinguals).

Hypothesis C was supported. Part C(a) of the hypothesis is that novel expressions (of either type) are less likely to be taken to make sense than conventional and literal ones are, and indeed we found them to be significantly less likely to be so. According to Hypothesis C(b), high-conforming novel expressions are more likely to be taken to make sense than low-conforming ones are, and we did find them significantly more likely to be so. The support for Hypothesis C(b) underscores the importance of addressing different types of novelty, and, in particular, provides some support for the distinctive way we have done this.

Hypothesis D(a) is that people take less time to assess whether literal or conventional metaphorical expressions make sense than novel metaphorical ones (whether low- or high-conforming). Hypothesis D(b) is that people will take different amounts of time to assess whether low-conforming and high-conforming novel expressions make sense. We found relevant significant timing differences for the adverb region and the answer region, though not in the noun region, as follows.

The results for the answer region supported Hypothesis D(b) and partially supported Hypothesis D(a). High-conforming novel metaphorical expressions attracted significantly slower meaningfulness judgements than all other conditions, but there were no other significant differences. This supported our expectation that

high-conforming novel metaphorical expressions would be processed at a different speed not just from the nonnovel conditions (an aspect of part (a)), but also more interestingly from low-conforming novel metaphors (part (b)). It also answers our question (in Section 2.3.2) about whether high-conforming novel metaphors would be processed more quickly or more slowly than low-conforming novel metaphors. However, the answer-region results did not support the aspect of Hypothesis D(a) where we predicted that low-conforming novel metaphors would be processed at a different speed from literal and conventional metaphorical expressions.

The results for the adverb region show partially supportive results for D(a), with high-conforming and low-conforming being significantly slower than conventional. However, the literal condition, surprisingly, did not differ significantly from the novel conditions. Hypothesis D(b) as applied to the adverb region was not supported, as there was no significant difference between the two novel conditions.

Despite the incompleteness of support for hypothesis D as applied to the adverb and answer regions, separately, the results for these regions taken together make it plausible that high-conforming novel expressions are more effort-involving than low-conforming ones, and that, in line with previous literature (see Section 2.3.1), novel expressions of both our types are more effort-involving than both literal and conventional expressions. This is promising for future research aimed at strengthening this plausibility, while raising specific issues needing further exploration, such as the speed of assessing the meaningfulness of low-conforming expressions versus literal and conventional ones.

As noted, in the answer region, high-conforming novel metaphors were processed significantly more slowly than low-conforming novel metaphors. One might have expected high-conforming novel metaphors to be faster to process, because they should not require new bridges (mappings, superordinate categories etc.), unusual within-source inferencing or other connection-following, or entertainment of special contexts. However, that only suggests a speed advantage over low-conforming expressions when rich comprehension is achieved in low-conforming as well as high-conforming cases. As suggested in the discussion of Hypothesis D in Section 2.3.2, many participants may have opted for less-rich comprehension in low-conforming cases, reducing or omitting work on new bridges, unusual within-source connection-following or special contexts. Indeed, many participants may have quickly rejected some low-conforming expressions as meaningless, rather than being more attentive and seeking rich meaning. The meaningfulness results show that participants took the pairs to made sense less often in the low-conforming cases.

5.2. The issue of condition/group interaction

Our results show no statistical interaction between metaphoricity condition and membership of the monolingual or multilingual group, whether as regards the polarity of meaningfulness judgements or the time taken to make them or read the sentences, despite the pattern of YES/NO answers in Table 6. Indeed, we had no clear basis on which to hypothesise an interaction. Our main interest as regards differences between multilinguals and monolinguals was to establish that they exhibit both some difference in the low-conforming condition and some difference in the high-conforming condition, rather than to show that such differences on individual conditions did or did not themselves differ from each other.

Nevertheless, having observed that multilinguals and monolinguals performed very similarly to each other in both the literal and conventional conditions, while performing differently in the novel conditions, it would be interesting to see whether a future, larger study would show a statistically significant condition/group interaction, at least involving the nonnovel/novel difference, if not the difference between the two novel conditions. In our study, the significant contrasts on novel conditions were presumably not strong enough to drive an interaction effect towards significance, due to an extremely striking similarity between the literal and conventional conditions.

An interaction of mono/multilingualism with the two types of novelty might arise if knowing additional languages, in general, or knowing specific additional languages, has an effect on differences in processing high-conforming versus low-conforming expressions (with high/low conformity here still being from the English viewpoint). For instance, it could make the person familiar with more metaphoric bridges (mappings, superordinate categories etc.). This would probably make more of a difference to low-conforming expressions that could now benefit from these bridges than to high-conforming expressions. It would effectively turn some English-wise low-conforming expressions into high-conforming ones from a certain other language's viewpoint. Nevertheless, for a study to reveal this, it would need to be systematic about, for instance, which additional languages multilinguals know, and it would need to include some expressions that engage bridges that are peculiar to the additional languages.

5.3. Some limitations of the current study

As with all linguistic materials in experiments, but especially when materials are novel and potentially puzzling, there is the complication of possible multiple interpretations. It was beyond the scope of the present study to filter out cases where participants constructed nonmetaphorical meanings for the pairs we classified as metaphorical, as we did not ask participants to state meanings that occurred to them. Nor, of course, could we check whether metaphorical meanings that they did discern exploited distinctive features of the adjectives' meanings, as was required for richest comprehension.

As noted in Section 3's preamble, our main reason for not asking participants to provide meanings for the pairs was that we did not assume that a participant's judgement that a sentence made sense would necessitate them having a particular meaning that was clear in their conscious mind and that therefore could be usefully reported during the experiment. A meaning could be largely absent from consciousness except in very broad strokes, or might be present but difficult to express; or there might be a range of alternative meanings that the participants would be hard pressed to choose from or to summarise. These possibilities are particularly salient for low-conforming expressions. For instance, it is difficult enough for us as metaphor researchers to state what a '*curved hope*' might be, even given considerable time to think about it, let alone for a non-metaphor researcher to be able to formulate a meaning during an experiment. An additional reason for not asking for meanings was that we wished to avoid the theoretically highly contentious process of judging whether stated meanings were reasonable, metaphorical ones.

Furthermore, in the case of our novel materials, it was inappropriate to use the common technique in metaphor studies of asking participants for lists of features or for their choices from lists of features or meanings provided by experimenters. This was because of the subtlety and structural complexity of possible interpretations, especially as different interpretations could have been divided largely by matters of degree. For instance, a *'curved hope'* (one of our low-conforming novel examples) could be taken to mean a hope that was somewhat more precisely defined and pleasing than a *'hazy hope'* (the corresponding high-conforming example), rather than differing from a hazy hope in a more black-and-white way.

A related limitation is that in our interpretation of the results in Section 5.1, we have taken YES answers on meaningfulness to mean the participants did have a genuine feeling of comprehension (whether or not because of consciously discerning clear, specific meanings), rather than feeling that they failed to comprehend but nevertheless wishing to save face by not appearing ignorant. However, note that if such face-saving were frequent enough to substantially invalidate our interpretations, we would have an interesting potential phenomenon of multilinguals engaging in it more than monolinguals, and of high-conforming expressions attracting it more often than low-conforming ones. Such possibilities are interesting targets for future research.

Because of the various limitations, and also for reasons of sample size, we were not able to make confident, separate analyses of reaction times for cases when participants said YES and cases where they said NO. It would naturally be beneficial in future work to try to surpass this limitation and the previously noted ones.

The findings from our study suggest that high-conforming novel phrases are comprehended more slowly than low-conforming ones. It would be intriguing in future work to filter out, in the low-conforming case, the YES responses that are based on less-rich, quick-to-derive meanings that do not exploit distinctive aspects of the phrases, thereby proceeding as if the expressions were high-conforming. If a speed advantage for low-conforming over high-conforming still remained after such filtering of YES answers, it might indicate that, for instance, the within-source inferencing or other connection-following and the use of familiar metaphoric bridges (mappings, superordinate categories etc.) in low-conforming cases tend to be less careful than in, or to hit obstacles earlier than in, actual high-conforming cases.

5.4. Further general discussion: cognitive flexibility and multilinguals

Our experimental results indicate that multilinguals tend to accept novel metaphors more than monolinguals do, which is in line with previous research on the 'cognitive flexibility' of multilinguals. Although we acknowledge the controversial nature of this claimed flexibility (Lehtonen et al., 2018; Paap & Greenberg, 2013; Papageorgiou et al., 2019), our findings suggest that differences in the way monolinguals and multilinguals process linguistic input exist, and need to be investigated, analysed and interpreted with different methods and tools.

Although the vague and generic labels 'cognitive advantage' or 'cognitive flexibility' may be easily misinterpreted, more fine-grained labels and specific effects of multilingualism on language processing remain an open and fertile field of investigation. Our current contribution goes in this direction, suggesting a specific sector in which monolinguals and multilinguals appear to perform in different ways, that is, a

possible linguistic flexibility: namely, in comprehending two different types of novel metaphors. This suggests that, if there is a cognitive advantage to multilingualism, it may relate more closely to divergent thinking (Runco & Acar, 2012) than to other kinds of flexibility that may be less closely related to novel metaphor comprehension.

5.5. Further general discussion: novelty and user relativity

There are a number of things to note concerning our approach to metaphor novelty. (We should stress that we continue with this article's focus on what novelty amounts to from the hearer's point of view. See Hidalgo-Downing (2020) for an overview of metaphoric creativity and novelty, one that encompasses speakers and hearers, context-sensitivity and multimodality.)[15] On the one hand, we have downplayed the role of lexical novelty: we have presumed in our study that an expression is, as a practical matter, more *likely* to be novel in our sense if it has only appeared rarely in corpora, but lexical novelty is not in fact part of our conception of metaphoric novelty. Moreover, in principle, an expression could have been frequently used with a particular meaning, but this meaning nevertheless may not have become entrenched in the sense of being simply retrievable from memory. Equally, in principle, it could happen that a very *infrequently* used expression could have a metaphorical meaning that becomes quickly entrenched for some special reason.

Furthermore, there is a hearer-relativity issue whereby different L1 hearers of a language may approach a given metaphorical expression in different ways. This is ultimately why we have defined our types of novelty on the basis of how hearers might process the expressions at hand, rather than trying to define them in a more traditional, hearer-neutral way.

The characterisations of our types of novelty are not strongly based on any specific theory of metaphor, but only require a very general notion of bridges between subject matters, together with the notion of drawing inferences within the source subject matter or following other sorts of connection within it, and the possibility that a hearer creatively (though perhaps unconsciously) entertains a special context when no helpful context is provided that points to a specific meaning. At the same time, our approach emphasises the point that the particular, *detailed* way in which an expression is novel as metaphor is ultimately a theory-relative issue. For instance, one theory might argue that a metaphor is novel because it requires new bridge construction, whereas another might say that a metaphor is novel because it requires unusual within-source inferencing, and a yet a third theory might say it requires both. In addition, these matters are, as we have already pointed out, language-relative. For example, literal translation of a metaphorical expression from one language to another can change whether it is metaphorically novel or not, or the precise way in which it is novel, partly because different bridges between source and target might be familiar to hearers. However, this issue itself interacts with what metaphor theory is postulated, because, for instance, the less

---

[15]Note that Hidalgo-Downing there lists 'novel metaphors' separately from 'creative elaborations of conventional metaphors' and 'elaborating and expanding source domains'. We take her to be using 'novel metaphors' to mean new ways of putting target and source subjects in correspondence – new bridges in our terms. Her other two categories are also types of novelty in our own view.

socio-culturally specific the types of bridge proposed by a given theory are, the less they might tend to vary between languages.

Our approach may shed new light on the different ways in which we might go beyond theories such as the career of metaphor theory as in Bowdle & Gentner (2005). In particular, this theory suggests a relatively straightforward gradation between conventional and novel metaphors, reflecting two different processing strategies: metaphor comprehension by means of construction of cross-domain mappings from scratch between source and target domains for the more novel metaphors, and metaphor comprehension by means essentially of the categorisation approach (if mere retrieval of meaning is not adequate) for the more conventional metaphors. Such a clear-cut distinction between metaphor types and processing strategies may need to be abandoned in favour of a more sophisticated set of distinctions and a rich array of different mixes of types of processing. Our distinction between different types of novelty addresses the fact that different levels of richness can be found for a given metaphorical expression, according in part to different extents to which the hearer exploits or ignores distinctive aspects of source-side concepts raised. Such a tendency to exploit or to ignore these aspects affects the extent to which the hearer on a specific comprehension occasion needs to, is inclined to, or is able to conjure up new mappings or other types of bridge between source and target, perform unusual within-source inferencing or other connection-following or entertain special contexts. These factors in turn can interact with each other. We therefore assume that one cannot distinguish types of novelty on any simple ground such as the binary distinction between requiring or not requiring new bridges, or measuring degree of novelty on any simple, unitary basis.

Despite these complications, we found that the high/low-conformity distinction did make a significant difference to meaningfulness judgements and the timing of comprehension attempts. A fertile area for further research would be to try to tease apart the different roles here of familiar-bridge availability, of possibilities for within-source inferencing or other connection-following, and of entertainment of special contexts. Moreover, in reality, there is a spectrum of degrees of conformity between what we have called high and low conformity, so it would be beneficial to investigate the dependence of timings and meaningfulness judgments on the whole range of degrees. How to measure the degrees would itself need to be a focus in such work.

Although we have sought to be as theory-neutral as possible in our study, our approach does chime especially well with potential metaphor comprehension theories that are eclectic and flexible in that: existing bridges can be used as they are, or refined when necessary; new bridges can be created when necessary; bridges may be of different types (e.g., mappings or superordinate categories) for different purposes (unless it can be shown that just one type works well under all circumstances); within-source connections of various types can be followed and the mixing and ordering of these various types of processing is flexible, dynamically arising and highly responsive to circumstances such as the context in which the discourse takes place, the linguistic context surrounding the expression currently being comprehended, and the purposes, interests, world/language experience and condition (mood, level of interest, alertness, distractedness, intellectual capacity etc.) of the hearer. From this point of view, the most promising metaphor theories are those that embed the question of metaphor within the question of semantic/pragmatic processing more broadly, or indeed cognition as a whole, such as

blending theory (Fauconnier & Turner, 1998, 2008), the cognitive-operation approach of Ruiz de Mendoza and colleagues (Ruiz de Mendoza, 2020; Ruiz de Mendoza & Galera, 2014), proposals within Relevance Theory, especially when incorporating inference about the literal, source-side scenario suggested by the expression at hand (Carston & Wearing, 2011) and the view of metaphor as dynamic/ecological performance (Gibbs Jr., 2019).

We now more explicitly address hearer relativity. It has long been generally recognised that the more novel a metaphor is, the more open-ended is the range of meanings that the hearer can justifiably and relevantly extract, and that different hearers and different circumstances can favour different meanings. In particular, individual differences in cognitive ability can have an effect on metaphor comprehension – see especially Stamenković, Ichien, & Holyoak (2020), whose study concentrated on literary metaphors, ones thus tending to the novel, such as *'Nerves after a quarrel are frozen leaves in winter'*. These authors note, however, that relatively few studies have explored this matter of individual ability differences. Relatedly, we argue that the hearer relativity of metaphoricity in general, and of conventional and novel metaphoricity in particular, needs to be more explicitly recognised when talking about conventional and novel metaphors from the hearer point of view. (Compare the comment by Hidalgo-Downing, 2020, that a linguistic expression will not be metaphoric or creative per se, but instead this will depend on how and in what context the expression is used.) It is not that, objectively, some expressions are or are not metaphorical, or are or are not conventional as metaphor or are or are not of this or that type of novelty as metaphor. Rather, the central point is that under particular circumstances, hearers process expressions in particular metaphoric ways, and these can involve (to differing extents) lexical retrieval, the creation or activation of old or new bridges, old or new within-source inferencing, and so forth. From this, as a practically useful but nevertheless crude and derivative abstraction, one can say that some expressions are conventional metaphorical ones because they have been frequently used with a particular fixed meaning (or, more probably in fact, with a cluster of highly context-sensitive but closely related meanings). However, such expressions might at any point be treated by a hearer in a way that partially or entirely ignores such meanings, and seeks meaning(s) afresh, although this does not preclude the hearer from landing on a meaning similar to established ones. Going back to the discussion in Section 2.2.1, such potential divergence from established meanings, or doing something more than just retrieving an established meaning, *may* normally occur due to pressure arising from an unusual linguistic or world context or accompanying multimodal communication, but equally may be a more general phenomenon. Or, it could sometimes be just because of the particular hearer's current condition or even random whim. On the other hand, an expression that has rarely before been used metaphorically and that potentially has rich meaning could be validly treated by a given hearer in a relatively crude way, and distinctive aspects of possible meaning may be ignored.

Therefore, categories such as 'conventional', 'novel' and 'creative' metaphors could usefully be renamed as 'conventionally amenable', 'novelty-enabling' or 'creativity-inspiring' or some such, when we are considering the hearer side of communication. 'High conforming' and 'low conforming' could more accurately be renamed as 'closely conformable' and 'distantly conformable'. It is, in the end, at best a preliminary and approximate venture to conduct experiments exploring hearers' treatment of novel or conventional expressions as if these categories had specifiable

nonhearer-relative natures that govern how hearers comprehend them. Instead, the direction of governing is more the opposite. There are different types of metaphoric processing available to hearers in a circumstance-specific way, and the way these are used on a given occasion governs the designation of a particular hearer's processing as involving metaphoricity or not, novelty or not and a particular type of novelty or not, on that occasion. The patterns of such processing across different hearers and over time govern the assignment of heuristically useful, but over-simplified and derivative, labels, such as conventional, novel etc., to metaphorical expressions themselves.

6. Conclusion

This study looked at two distinct, but connected, research areas: the nature of metaphorical novelty, and the ability of multilinguals as contrasted with monolinguals to comprehend novel metaphorical language. Our experimental results indicate that multilinguals tend to accept novel metaphors more than monolinguals do, which is in line with previous empirical indications of other sorts of 'cognitive flexibility' of multilinguals, and thus tending to show that multilinguals do have some such flexibility advantage despite some controversy on this point.

We addressed novel metaphor of two types, carved out from the opposing ends of what is in reality a complex spectrum of novelty. One of the types was high-conforming novelty, where distinctive meaning based on distinctive features of the source items can straightforwardly be found without having to invent new mappings or other bridges between source and target subject matters, and without having to entertain special contexts or do unusual inferencing or other following of connections within the source subject matter. The other type was low-conforming novelty, where rich comprehension requires marked use of such types of processing. This particular way of distinguishing types of novelty may be of value to other researchers. Our results also suggest that, although people find high-conforming novel metaphorical phrases easier to comprehend than low-conforming ones, they spend more time on the former. This may suggest that in the presence of low-conforming novelty, where special processing is needed, people tend to give up quickly.

The language-cognition relationship may be influenced by the nature and on-the-fly condition of the individual participants, the historical and geographical context, the existence of familiar source/target bridges (mappings etc.) in the cognitive underpinnings of the other languages that they speak and other contingent factors. Indeed, given the complex and dynamic nature of the relationship between language and cognition, much further research, both theoretical and empirical, is suggested by our study.

References

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Bambini, V., Canal, P., Resta, D. & Grimaldi, M. (2019). Time course and neurophysiological underpinnings of metaphor in literary context. *Discourse Processes* 56(1), 77–97.

Barnden, J. A. (2015). Open-ended elaborations in creative metaphor. In T. R. Besold, M. Schorlemmer & A. Smaill (eds), *Computational creativity research: Towards creative machines*, 217–42. Atlanta: Atlantis Press.

Barnden, J. A. (2016). Communicating flexibly with metaphor: A complex of strengthening, elaboration, replacement, compounding and unrealism. *Review of Cognitive Linguistics. Published under the Auspices of the Spanish Cognitive Linguistics Association* 14(2), 442–73.

Barnden, J. A. (2020). Metaphor thoughtfully. In A. Baicchi (ed.), *Figurative meaning construction in thought and language*, 13–43. Amsterdam: John Benjamins.

Barnden, J.A. & Lee, M.G. (2001). Application of the ATT-Meta metaphor-understanding system to an example of the metaphorical view of MIND PARTS AS PERSONS. *Technical Report CSRP-01-09, School of Computer Science*, The University of Birmingham, U.K., 2001.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarello (eds), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*, 255–84. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bialystok, E. (2001a). *Bilingualism in development: Language, literacy, and cognition*. Cambridge: Cambridge University Press.

Bialystok, E. (2001b). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics* 21, 169–81.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 65(4), 229–35.

Bialystok, E., Craik, F. I., Green, D. W. & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest* 10(3), 89–129.

Bialystok, E. & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science* 7(3), 325–39.

Bialystok, E., Poarch, G., Luo, L. & Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging* 29(3), 696–705.

Birdsell, B. (2018a). Conceptual wandering and novelty seeking: Creative metaphor production in an L1 and L2. *Journal of Cognitive Science* 19(1), 35–51.

Birdsell, B. (2018b). Creative metaphor production in a first and second language and the role of creativity. Unpublished PhD dissertation, University of Birmingham.

Blasko, D. G. & Briihl, D. S. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol* 12(4), 261–85.

Blasko, D. G. & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(2), 295–308.

Bowdle, B. F. & Gentner, D. (2005). The career of metaphor. *Psychological Review* 112(1), 193–216.

Butler, Y. G. (2013). Bilingualism/multilingualism and second-language acquisition. In T. K. Bhatia & W. C. Ritchie (eds.) *The handbook of bilingualism and multilingualism*, 109–36. Hoboken, NJ: John Wiley & Sons.

Cacciari, C., Bolognini, N., Senna, I., Pellicciari, M. C., Miniussi, C. & Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: A TMS study. *Brain and Language* 119(3), 149–57.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A. & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *Neuroimage* 59(4), 3212–21.

Carston, R. & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition* 3(2), 283–312.

Casasanto, D. & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistics Vanguard* 1 (1), 327–37.

Chen, S. Y. & Husband, E. M. (2018). Contradictory (forward) lifetime effects and the non-future tense in Mandarin Chinese. *Proceedings of the Linguistic Society of America* 3(6), 1–14.

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, 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.

Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.

de Bruin, A., Bak, T. H. & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language* 85, 15–26.

de Bruin, A., Dick, A. S. & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language* 2(4), 433–51.

de Bruin, A., Treccani, B. & Della Sala, S. (2015a). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science* 26(1), 99–107.

de Bruin, A., Treccani, B. & Della Sala, S. (2015b). The connection is in the data: We should consider them all. *Psychological Science* 26(6), 947–49.

Fauconnier, G. & Turner, M. (1998). Conceptual integration networks. *Cognitive Science* 22(2), 133–87.

Fauconnier, G. & Turner, M. (2008). Rethinking metaphor. In R. W. Gibbs Jr. (ed.), *The Cambridge handbook of metaphor and thought*, 53–66. Cambridge: Cambridge University Press.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development* 59(1), 47–59. https://doi.org/10.2307/1130388

Gentner, D. & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language* 37(3), 331–55.

Gibbs Jr., R. W. (2019). Metaphor as dynamical–ecological performance. *Metaphor and Symbol* 34(1), 33–44.

Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N. & Zur, A. (2004). Weapons of mass distraction: Optimal innovation and pleasure ratings. *Metaphor and Symbol* 19(2), 115–41. https://doi.org/10.1207/s15327868ms1902_2

Glucksberg, S. (1998). Understanding metaphors. *Current Directions in Psychological Science* 7(2), 39–43.

Glucksberg, S. (2001). *Understanding figurative language: From metaphor to idioms*. Oxford: Oxford University Press. Retrieved from https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195111095.001.0001/acprof-9780195111095.

Grady, J. E. (1997). Theories are buildings revisited. *Cognitive Linguistics (Includes Cognitive Linguistic Bibliography)* 8(4), 267–90.

Hampe, B. (2017). Embodiment and discourse: Dimensions and dynamics of contemporary metaphor theory. In B. Hampe (ed.), *Metaphor: embodied cognition and discourse*, 3–23. Cambridge: Cambridge University Press.

Hidalgo-Downing, L. (2020). Introduction: Towards an integrated framework for the analysis of metaphor and creativity in discourse. In L. Hidalgo-Downing & B. Kraljevic-Mujic (eds), *Performing metaphoric creativity across modes and contexts*, 1–17. Amsterdam: John Benjamins.

Holyoak, K. J. & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin* 144(6), 641–71.

Jacques, S. & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology* 20(3), 573–91.

Jarvis, S. (2003). Probing the effects of the L2 on the L1: A case study. *Effects of the Second Language on the First* 3, 81–102.

Kecskés, I. & Papp, T. (2000). Metaphorical competence in trilingual language production. In J. Cenoz & U. Jessner (eds), *English in Europe: The acquisition of a third language*, 99–120. Clevedon: Multilingual Matters.

Kunzetsova, A., Brockhoff, P. & Christensen, R. (2017). lmerTest package: Tests in linear mixed effect models. *Journal of Statistical Software* 82, 1–26.

Lai, V. T. & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language* 127(3), 484–96.

Lai, V. T., Curran, T. & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research* 1284, 145–55.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*, 1st ed. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Lakoff, G. & Turner, M. (2009). *More than cool reason: A field guide to poetic metaphor*. Chicago: University of Chicago press.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A. & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin* 144(4), 394–425.

Lenth, R. (2021). emmeans: Estimated marginal means, aka least-squares means. R package version 1.6.0. (2021).

Littlemore, J. (2001). Metaphoric competence: A language learning strength of students with a holistic cognitive style? *Tesol Quarterly* 35(3), 459–91.

Littlemore, J. (2010). Metaphoric competence in the first and second language. In M. Putz & L. Sicola (eds), *Cognitive processing in second language acquisition*. Converging Evidence in Language and Communication Research, Vol. 13, 293–316. Amsterdam: John Benjamins.

Littlemore, J. & Low, G. (2006a). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics* 27(2), 268–94.

Littlemore, J. & Low, G. D. (2006b). *Figurative thinking and foreign language learning*. Palgrave Macmillan.

Littlemore, J., Pérez-Sobrino, P., Houghton, D., Shi, J. & Winter, B. (2018). What makes a good metaphor? A cross-cultural study of computer-generated metaphor appreciation. *Metaphor and Symbol* 33(2), 101–22. https://doi.org/10.1080/10926488.2018.1434944

Martin-Rhee, M. M. & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism* 11(1), 81–93.

Mashal, N. & Faust, M. (2009). Conventionalisation of novel metaphors: A shift in hemispheric asymmetry. *Laterality* 14(6), 573–89.

Müller, C. (2009). *Metaphors – Dead and alive, sleeping and waking: A dynamic view*. Chicago: University of Chicago Press. Retrieved from https://www.degruyter.com/view/product/535157.

Murphy, V. A. & Pine, K. J. (2003). L2 influence on L1 linguistic representations. In V. Cook (ed.), *Effects of the second language on the first*, 142–67. Bristol and Blue Ridge Summit: Multilingual Matters.

Paap, K. R. & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* 66(2), 232–58.

Papageorgiou, A., Bright, P., Periche Tomas, E. & Filippi, R. (2019). Evidence against a cognitive advantage in the older bilingual population. *Quarterly Journal of Experimental Psychology* 72(6), 1354–63.

Ruiz de Mendoza, F. J. (2020). Understanding figures of speech: Dependency relations and organizational patterns. *Language and Communication* 71, 16–38.

Ruiz de Mendoza, F. J. & Galera, A. (2014). *Cognitive modeling. A linguistic perspective*. Amsterdam and Philadelphia: John Benjamins.

Runco, M. A. & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal* 24(1), 66–75.

Rutter, B., Kröger, S., Hill, H., Windmann, S., Hermann, C. & Abraham, A. (2012). Can clouds dance? Part 2: An ERP investigation of passive conceptual expansion. *Brain and Cognition* 80(3), 301–10.

Sabourin, L., Brien, C. & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition* 17(3), 542–55.

Sabourin, L., Leclerc, J.-C., Lapierre, M., Burkholder, M. & Brien, C. (2016). *The language background questionnaire in L2 research: Teasing apart the variables.* Paper presented at the Proceedings of the 2016 Annual Conference of the Canadian Linguistic Association, Calgary, AB.

Sperber, D. & Wilson, D. (2008). A deflationary account of metaphors. In R. Gibbs (ed.) *The Cambridge handbook of metaphor and thought*, 84–105. Cambridge: Cambridge University Press.

Stamenković, D., Ichien, N. & Holyoak, K. J. (2020). Individual differences in comprehension of contextualized metaphors. *Metaphor and Symbol* 35(4), 285–301.

Tomoschuk, B., Ferreira, V. S. & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition* 22(3), 516–36.

Werkmann Horvat, A., Bolognesi, M. & Kohl, K. (2021). Creativity is a toaster: Experimental evidence on how multilinguals process novel metaphors. *Applied Linguistics* 42(5), 823–47.

Werkmann Horvat, A., Bolognesi, M. & Lahiri, A. (2021). Processing of literal and metaphorical meanings in polysemous verbs: An experiment and its methodological implications. *Journal of Pragmatics* 171, 131–46.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. London: Routledge.

Wolff, P. & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science* 35(8), 1456–88.

Yelland, G. W., Pollard, J. & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics* 14(4), 423–44.