Individual Differences in Comprehension of Contextualized Metaphors

Dušan Stamenković, Nicholas Ichien & Keith J. Holyoak

To cite this article: Dušan Stamenković, Nicholas Ichien & Keith J. Holyoak (2020) Individual Differences in Comprehension of Contextualized Metaphors, Metaphor and Symbol, 35:4, 285-301

To link to this article: https://doi.org/10.1080/10926488.2020.1821203

Published online: 28 Oct 2020.
Individual Differences in Comprehension of Contextualized Metaphors

Dušan Stamenković\textsuperscript{a}, Nicholas Ichien\textsuperscript{b}, and Keith J. Holyoak\textsuperscript{a,b}

\textsuperscript{a}University of Niš; \textsuperscript{b}University of California, Los Angeles

\section*{ABSTRACT}
We report a study examining the role of linguistic context in modulating the influences of individual differences in fluid and crystalized intelligence on comprehension of literary metaphors. Three conditions were compared: no context, metaphor-congruent context, and literal-congruent context. Relative to the baseline no-context condition, the metaphor-congruent context facilitated comprehension of the metaphorical meaning whereas the literal-congruent context impaired it. Measures of fluid and crystalized intelligence both made separable contributions to predicting metaphor comprehension. The metaphor-congruent context selectively increased the contribution of crystalized verbal intelligence. These findings support the hypothesis that a supportive linguistic context encourages use of semantic integration in interpreting metaphors.

\section*{Introduction}
A metaphor aims to illuminate a thing by describing it in terms of another thing quite unlike it. Consider this example from Vincent van Gogh (1882): “Conscience is a man’s compass, and though the needle sometimes deviates, though one often perceives irregularities when directing one’s course by it, one must try to follow its direction.” Here the target concept (what is being talked about) is a conscience, and the source (used to characterize the target) is a compass.\textsuperscript{1} Though a conscience is not a literal compass, the parallels between the functions of a compass as a useful if imperfect geographical guide help to make a point about the functions of a conscience as a moral guide for navigating the “path” of life. More generally, a metaphor links a target that is unfamiliar, abstract, or poorly understood to a source that is familiar, concrete, or better understood. In addition to attribution of salient features of the source domain to the target, metaphors typically result in the attribution of emergent properties that are not necessarily associated with the source and target in isolation; moreover, these properties are not always derivable by rules of semantic composition (Wilson & Carston, 2006). Metaphor is widely acknowledged to be a core manifestation of human language and thought (e.g., Lakoff & Johnson, 1980). But despite decades of research, no consensus has been reached about the psychological mechanisms that support comprehension of metaphors (for a recent review, see Holyoak & Stamenković, 2018). Although most metaphors occur within a particular linguistic or pragmatic context, relatively little work has examined the role played by context in metaphor processing. In addition, although evidence suggests that people differ in their ability to process metaphors, the nature of these individual differences remains unclear. The present paper is, to the best of our knowledge, the first to examine individual differences in the comprehension of contextualized metaphors.

\section*{CONTACT} Dušan Stamenković \textsuperscript{a} dustan.stamenkovic@filfak.ni.ac.rs Faculty of Philosophy & Center for Cognitive Sciences, University of Niš, Niš 18000, Serbia

\section*{Note:} The research data for this paper is available in Mendeley Data: https://dx.doi.org/10.17632/gykynsmmvcc.1

\textsuperscript{1}The terms target and source are commonly used in work on analogy, as well as in cognitive linguistics (Lakoff & Turner, 1989). In linguistics, the corresponding terms tenor and vehicle are often used.

© 2020 Taylor & Francis Group, LLC
Models of metaphor processing

The capacity to produce and comprehend metaphors is a distinctively human ability that depends on the interplay between language and thought. For over half a century, psychologists have attempted to develop models of metaphor processing based on some mix of general reasoning and semantic knowledge. Two broad proposals have dominated theoretical discussions. A view that traces back to Aristotle assumes that metaphor comprehension requires analogical reasoning to relate the target to the source (Gentner & Clement, 1988; Tourangeau & Sternberg, 1981, 1982; Trick & Katz, 1986). Black’s (1962) interaction theory of metaphor also assumed that metaphor is based at least in part on analogy, postulating “interactions between two systems, grounded in analogies of structure (partly created, partly discovered)” (Black, 1977/1993, p. 39). Although analogies can be posed verbally, analogical reasoning can also be applied to formal, nonverbal content such as geometric patterns (e.g., Mulholland, Pellegrino, & Glaser, 1980) and mathematical problems (e.g., Novick & Holyoak, 1991). The dominant analogy-based model of metaphor (Bowdle & Gentner, 2005; Gentner & Bowdle, 2008) claims that metaphor comprehension (at least for novel metaphors) involves the application of a domain-general process for finding systematic mappings between elements of the source and target situations (e.g., for van Gogh’s metaphor, emotions triggered by one’s conscience may be mapped to the indicator on a compass). Given that analogical mapping appears to depend on active manipulation of complex propositional structures, and consideration of multiple alternative hypotheses, this process is likely to place a high load on working memory and executive functions such as attentional control (e.g., Cho, Holyoak, & Cannon, 2007; Waltz, Lau, Grewal, & Holyoak, 2000).

A second general approach to metaphor comprehension focuses on potential mechanisms internal to the linguistic system. A number of theorists have proposed that metaphor comprehension is based on processes of semantic integration continuous with those involved in literal language comprehension. In general terms, semantic integration is the incremental construction of the meaning of a message from the meanings of individual words, guided by both syntactic structure and the pragmatic context of the utterance (e.g., Bransford & Franks, 1971; Kintsch & Mangalath, 2011). Semantic integration is a broad concept, which can be viewed as continuous with the construction of situation models during text comprehension (Kintsch & van Dijk, 1978). It encompasses more specific processes such as conceptual combination (e.g., integrating the meanings of the words in a noun-noun compound; Wisniewski & Gentner, 1991) and predication (e.g., integrating the meanings of the words in a noun-verb expression; Kintsch, 2001). The well-known categorization theory of metaphor (Glucksberg & Keysar, 1990) can be viewed as a special case of predication that applies to sentences in the “is a” syntactic form. Whereas analogical reasoning is typically viewed as a domain-general process that operates on complex propositional structures held in working memory (Holyoak, 2012), semantic integration appears to be based on processes less dependent on executive functions and more dependent on semantic processing (e.g., operations on vectors of semantic features) that operate at the level of lexical semantics (Keane & Costello, 2001).

A particularly influential proposal, introduced by Glucksberg and Keysar (1990) and extended by others (e.g., Gernsbacher, Keysar, Robertson, & Werner, 2001; Glucksberg & Haught, 2006; Jones & Estes, 2005, 2006; McGlone & Manfredi, 2001), claims that metaphors are interpreted as categorization statements, in which the meaning of the source is contextually adjusted to create a novel category. For example, in van Gogh’s metaphor, the literal meaning of compass as a navigational instrument might be modified to form a more abstract category of “directional guide”. As noted above, metaphor-as-categorization can be modeled as a case of predication (Kintsch, 2000, 2001; Kintsch & Bowles, 2002). In Kintsch’s predication model, a process of spreading activation activates concepts that link the source (expressed as the object of a linking verb, e.g., “is a man’s compass”) and target (e.g., “conscience”). The semantic representation of the source concept is then modified to create a new semantic representation, which in turn shades the meaning of the target.

The relative contributions of analogical reasoning and semantic integration to metaphor processing remain uncertain. It has been suggested that novel metaphors are interpreted using analogy, whereas
more familiar metaphors are processed as category statements (Bowdle & Gentner, 2005). However, neural evidence suggests that metaphor processing does not primarily depend on brain regions associated with analogical reasoning. Neuroimaging studies have found that metaphor comprehension typically involves activation of areas within the temporal cortex also associated with comprehension of literal language (for meta-analyses see Bohn, Altmann, & Jacobs, 2012; Rapp, Mutschler, & Erb, 2012; Vartanian, 2012), with additional right-hemisphere involvement in the case of novel metaphors (e.g., Arzouan, Goldstein, & Faust, 2007). By contrast, the type of complex relational integration involved in analogical reasoning is associated with activation of the left rostrolateral prefrontal cortex (for meta-analyses see Vartanian, 2012; Hobeika, Diard-Detoeuf, Garcin, Levy, & Volle, 2016; see also Bunge, Helskog, & Wendelken, 2009). This area typically is not activated during metaphor tasks (but see Prat, Mason, & Just, 2012), even for novel metaphors (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012). In another dissociation, people with autism tend to show deficits in metaphor comprehension (Kalandadze, Bambina, & Næss, 2019; Morsanyi, Stamenković, & Holyoak, 2020a), even though analogical reasoning is an area in which they exhibit relative strength (Morsanyi, Stamenković, & Holyoak, 2020b).

A significant caveat on any conclusions about the processes underlying metaphor comprehension is that the novel metaphors used in research have generally been constructed by psychologists. Relative to familiar metaphors, these novel metaphors have tended to be less apt (i.e., the source describes the target less fully and accurately; Bowdle & Gentner, 2005; Jones & Estes, 2006). It is possible that literary metaphors – those created by poets and other authors – have different qualities than those constructed by psycholinguists and cognitive psychologists. A large set of literary metaphors was collected and normed by Katz, Paivio, Marschark, and Clark (1988). This item set has been extensively analyzed by Jacobs and Kinder (2017, 2018) using machine-learning algorithms and other quantitative methods. Those metaphors in the Katz et al. collection that were created by poets differ in many ways from those created by metaphor researchers. Although these differences are often subtle, machine-learning algorithms are able to distinguish literary from nonliterary metaphors with high accuracy. Jacobs and Kinder (2018) found that qualities distinguishing literary metaphors rated high in goodness include high surprisal (a statistical measure of the unexpectedness of words), relative dissimilarity of source and target concepts, the combination of concrete words with relatively complex grammar and high lexical diversity, and extra difficulty in comprehending the metaphorical meaning (see also Holyoak, 2019; Lakoff & Turner, 1989). These properties collectively suggest that good literary metaphors are high in cognitive complexity, which may be more likely to elicit analogical reasoning.

**Comprehension of contextualized metaphors**

A limitation of research on metaphor comprehension is that most studies have examined processing of relatively simple metaphors (typically in the nominal syntactic form, `<noun> is <noun>`) in the absence of a broader linguistic or pragmatic context. Much less is known about how metaphors are interpreted when a context is provided (Kövecses, 2015, pp. x–xi), especially in the case of literary metaphors. Note that van Gogh did not simply say, “Conscience is a man’s compass;” rather, he went on to elaborate on some key correspondences. Moreover, in the passage *preceding* the metaphor, he wrote: "I think it one’s duty to try to do the right thing, even knowing that one cannot go through life without making mistakes, without regret or sorrow. Somewhere I read, Some good must come by clinging to the right.” The metaphor was thus introduced as the culmination of a theme already established by the prior context. Gibbs and Gerrig (1989) observed that we rarely encounter metaphors in isolation, but rather in the context of a conversation, or a piece of poetry or prose. Metaphors are generally part of some larger discourse context that provides a conceptual framework that helps a reasoner take what *is said* by a metaphorical expression and generate an interpretation of what *is meant* by that expression. Metaphor is considered to be a deeply context-sensitive communicative phenomenon (Camp, 2006).
Based on the results of several psycholinguistic studies, Carston (2012) argued that metaphorical interpretations are derived automatically when the context makes them accessible, and that people do not take longer to comprehend metaphors than literal language when metaphors are properly contextualized. Experimental evidence has indeed established that a prior context congruent with a metaphorical interpretation indeed facilitates comprehension (Gerring & Healy, 1983; Gildea & Glucksberg, 1983; Inhoff, Lima, & Carroll, 1984; Ortony, Schallert, Reynolds, & Antos, 1978), whereas a prior context congruent with an alternative literal interpretation interferes with metaphor comprehension (McGlone & Manfredi, 2001). Shinjo and Myers (1987) found that contextual primes influence literal and metaphorical sentences in a similar fashion, and that relatively specific contexts are most effective. A recent neuroimaging study found no differences between the neural processing of metaphors and matched literal sentences when both were surrounded by a suitable context (Hartner et al., 2020). Providing a contextualized metaphorical setting can facilitate the formation of metaphor-based schemas, which are used to link elements within a text representation (Allbritton, 1995; Allbritton, McKoon, & Gerrig, 1995). Several other studies have confirmed the beneficial effect of congruent context on understanding figurative language (e.g., Gagné, Friedman, & Faries, 1996; Harris, 1976; Nayak & Gibbs, 1990; Pexman, Ferretti, & Katz, 2000; Rumelhart, 1979/1993; Thibodeau & Durgin, 2008; for a philosophical account of metaphor in context, see Stern, 2000). In general, longer, more specific, and more thematically congruent contexts provide the greatest benefit for metaphor comprehension. It is less clear how context exerts its effects (e.g., by facilitating either analogical reasoning or semantic integration, or by shifting the balance between these two general mechanisms). Given that a congruent linguistic context will tend to include words that share semantics with the metaphorical meaning, perhaps the most plausible hypothesis is that a congruent context facilitates semantic integration for the metaphorical meaning (Lemaire & Bianco, 2003).

**Individual differences in metaphor comprehension**

A relatively small number of studies have examined the impact of individual differences in cognitive abilities on metaphor comprehension (e.g., Chiappe & Chiappe, 2007; Kazmerski, Blasko, & Dessalegn, 2003; Olkonieni, Ranta, & Kaakinen, 2016; Trick & Katz, 1986). Based on classical theories of intelligence (Cattell, 1971), two partially separable factors likely to be relevant to metaphor processing are fluid and crystallized intelligence. Fluid intelligence (closely related to individual differences in executive functions) involves reasoning (often nonverbal) about novel problems detached from prior knowledge, and is important for success in explicit analogical reasoning (Holyoak, 2012). Crystallized intelligence involves reasoning (typically verbal) that draws upon prior knowledge, and hence is likely to impact semantic integration, which depends on knowledge of lexical semantics.

Metaphor comprehension seems likely to depend on both forms of intelligence, and several approaches to metaphor comprehension based on individual differences have confirmed this general hypothesis. Chiappe and Chiappe (2007) found that individuals who scored high on a working-memory test generated higher-quality interpretations of metaphors more quickly. In addition, measures of inhibitory control (based on Stroop interference and intrusion errors on a memory test) predicted metaphor processing (also see Pierce & Chiappe, 2008). Both working memory and inhibitory control are executive functions closely linked to fluid intelligence (Ackerman, Beier, & Boyle, 2005). In a production task, Chiappe and Chiappe (2007) found that measures of vocabulary knowledge and exposure to printed text (linked to crystallized intelligence) also predicted metaphor quality. Indeed, crystallized intelligence yielded somewhat higher correlations with metaphor interpretation and production than did measures of working memory. Overall, these findings suggest that both fluid and crystallized intelligence have an impact on metaphor processing. In addition, studies by Silvia and Beaty (2012; Beaty & Silvia, 2013) indicate that fluid intelligence is strongly related to creativity in a metaphor production task, while crystallized knowledge predicts the ability to generate conventional metaphors.
Stamenković, Ichien, and Holyoak (2019) took an individual-differences approach to examine the comprehension of metaphors drawn from literary as well as nonliterary sources. The literary metaphors were selected from the set collected and normed by Katz et al. (1988), and extensively analyzed by Jacobs and Kinder (2017, 2018) using machine-learning algorithms. Stamenković et al. measured fluid intelligence using a short form of the nonverbal Raven’s Progressive Matrices test (Arthur, Tubre, Paul, & Sanchez-Ku, 1999; Raven, 1938). Crystalized verbal intelligence was assessed using a novel Semantic Similarities Test (SST), in which word pairs are presented with the question, “How are the two concepts in each pair similar to one another?” Scores on the SST correlated well with the Vocabulary subscale of the Wechsler Adult Intelligence Scale, a standard measure of crystalized intelligence. For nonliterary metaphors, regression analyses indicated that only the measures of crystalized intelligence predicted unique variance in comprehension scores. In contrast, for literary metaphors measures of both fluid and crystalized intelligence made separable contributions to predicting comprehension. This pattern suggests that analogical reasoning (known to be strongly linked to fluid intelligence) may play an important role in comprehension only for relatively complex metaphors, whereas some form of semantic integration (more dependent on crystalized intelligence) influences metaphor comprehension across a broader range of metaphor types (nonliterary as well as literary).

**Overview of present study**

The present study sought to determine whether and how individual differences in fluid and crystalized intelligence impact the comprehension of literary metaphors when a prior linguistic context is provided. We sought to extend the approach of Stamenković et al. (2019), who only examined metaphors presented as isolated sentences, without a context. We focused on literary metaphors because our earlier findings indicated that for these cognitively complex metaphors, in the absence of a context both types of cognitive abilities contribute to prediction of comprehension.

Here we assessed comprehension of a metaphor (presented as a critical sentence) in three conditions: (1) no context (a baseline condition comparable to that tested by Stamenković et al., 2019), (2) metaphor-congruent context (preceding text supporting the metaphorical interpretation of the critical sentence), and (3) literal-congruent (preceding text focusing on target only, rather than supporting the metaphorical interpretation of the critical sentence). The third condition is similar to what previous researchers (Keysar, Shen, Glucksberg, & Horton, 2000; Thibodeau & Durgin, 2008) have termed a “no-mapping” or “non-metaphorical” scenario.

Based on previous work on context effects we predicted that, relative to the baseline no-context condition, a metaphor-congruent context would facilitate comprehension of the metaphorical meaning whereas a literal-congruent context would impair it. Our central aim was to assess how the context variations would influence the pattern of individual differences. Based on the previous findings of Stamenković et al. (2019) for literary metaphors, we expected measures of both fluid and crystalized intelligence to make separable contributions to predicting metaphor comprehension, at least in the no-context condition. We hypothesized that the metaphor-congruent context would encourage and facilitate semantic integration as the dominant strategy for interpreting the critical sentence, perhaps resulting in greater reliance on crystalized intelligence. It was less clear how the literal-congruent context would impact individual differences.

**Method**

**Participants**

Participants were 201 undergraduates at the University of California, Los Angeles (UCLA) (female = 152, male = 48, genderqueer = 1; mean age = 19.3) who received course credit. The study was approved, including informed consent procedures, by the UCLA Office of the Human Research Protection Program. This sample size is consistent with those used in similar studies
focusing on individual differences in metaphor comprehension (Stamenković et al., 2019), relational similarity (Ichien, Lu, & Holyoak, 2019), and relational reasoning (Gray & Holyoak, 2020). The participants were mainly native speakers of English (79%), with a minority of bilinguals who spoke English fluently (21%, self-assessed). We adopted the exclusion criteria based on those used by Stamenković et al. (2019) to identify carelessness or inattention on verbal tasks. Following these criteria, we did not analyze data from an additional 15 participants who either scored below 12 (max = 40) on the Semantic Similarities Test, below 10 (max = 40) on the metaphor test, or failed to complete the entire battery of tests. After these exclusions, a total of 65 participants were assigned to the no-context condition, 69 to the metaphor-congruent condition, and 67 to the literal-congruent condition.

**Design, materials, and procedure**

All participants completed four tasks in a fixed order. The first three were assessments of individual differences on cognitive tasks, and the fourth and final task involved metaphor comprehension. The task order was identical to that used by Stamenković et al. (2019, Study 3); the order was constant so as to minimize variability in the predictive tasks (as our focus was on individual differences). All tasks were administered to participants individually by computer. None of the tasks involved any sort of time pressure and participants could take breaks between tasks. The entire test session lasted approximately 50 minutes to an hour.

**Task 1: Raven’s Progressive Matrices (RPM).** A short version of the RPM (Arthur et al., 1999), adapted for computer administration using SuperLab software, was administered. Each matrix problem consists of a 3 × 3 array in which geometric forms vary systematically across rows and columns. The lower right cell is empty, and the task is to select the best completion from a set of eight options. The RPM is thus similar to geometric analogy problems, and is generally considered a central measure of fluid intelligence (Snow, Kyllonen, & Marshalek, 1984). In addition to predicting performance on psychometric analogy tests, scores on the RPM correlate with production of relational responses in a picture-mapping task (Vendetti, Wu, & Holyoak, 2014), spontaneous analogical transfer in a problem-solving task (Kubricht, Lu, & Holyoak, 2017), and similar tasks requiring relational reasoning (Gray & Holyoak, 2020). To the extent that explicit analogical reasoning is required to comprehend metaphors, the RPM should be a robust predictor of individual differences in metaphor comprehension.

**Task 2: Semantic Similarities Test (SST).** This test, introduced by Stamenković et al. (2019), provides a rapid assessment of crystalized verbal intelligence with face validity of relevance to metaphor comprehension. The SST is designed to measure participants’ ability to identify similarities between concepts expressed as pairs of single words, where the similarities vary in degree of abstraction. The test was designed to be similar to the Wechsler Adult Intelligence Scale (WAIS) Similarities subscale (a measure of verbal comprehension), but uses entirely different items, selected to span a broader range of similarities. The SST is a verbal test in which semantic knowledge of word meanings is critical. In addition to predicting comprehension of both literary and nonliterary metaphors (Stamenković et al., 2019), the SST has been found to predict performance on a battery of relational reasoning tasks (Gray & Holyoak, 2020).

**Task 3: Vocabulary Subtest of Wechsler Adult Intelligence Scale.** The third test was the Vocabulary subtest of the third version of the Wechsler Adult Intelligence Scale (WAIS-III), a standardized measure of crystalized intelligence consisting of 33 vocabulary items. Responses were assessed using the official scoring manual (Wechsler, 1997).

**Task 4: Literary metaphor comprehension.** The final task consisted of 20 literary metaphor comprehension items selected from a list of literary metaphors drawn from poetry anthologies by Katz et al. (1988), and rated high on their “goodness” scale. Each metaphor was in the form of a single critical sentence (e.g., *Nerves after a quarrel are frozen leaves in winter*; see Table 1). Their syntactic forms included nominal (*<noun> is <noun>*), nominal with an adjective modifier, and nominal with
Table 1. Context conditions and scoring key for an example of a literary metaphor: Nerves after a quarrel are frozen leaves in winter.

<table>
<thead>
<tr>
<th>Context Text</th>
<th>Scoring Key</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metaphor-congruent Condition</strong></td>
<td><strong>Literal-congruent condition</strong></td>
</tr>
<tr>
<td>We were fighting from early in the morning until lunch. Many awful things were said, and at some points I could not believe what was happening. Now that it has finally stopped, my whole body is shaking. I can't really think reasonably.</td>
<td>A quarrel is an angry argument or disagreement. There are people who fight all the time, and it's sometimes hard to understand why. It's simply part of who they are and what they are. There is no point in trying to understand such people.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Points</th>
<th>Sample responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 points (completely correct)</td>
<td>After a quarrel, it is difficult to do anything. A person may be deeply affected by arguments and confrontation. Nerves are preventing one from fully thinking about the situation.</td>
</tr>
<tr>
<td>1 point (partially correct)</td>
<td>Nerves leave you speechless after fighting. After arguing a person is shocked. Nerves after a fight can be stunning and make you feel paralyzed/numb.</td>
</tr>
</tbody>
</table>

a prepositional phrase (for detailed analyses of the properties of this metaphor set, see Jacobs & Kinder, 2018). All the literary metaphors used in the study are provided in Supplemental Online Materials, Appendix I.

The metaphors were presented one at a time on a computer display, in a different random order for each participant. Comprehension was assessed by an open-ended task requiring participants to type in their interpretation of each of the metaphorical statements. An initial example was provided. In the no-context condition, participants were simply instructed to “tell us how you understand the meaning” of each expression, responding by typing their answer in a box on the computer screen. For the two context conditions, the full passage (context followed by the critical sentence, which was underlined) was shown on the screen. Participants were instructed, “Your task is to read each text and the ending expression and tell us how you understand the meaning of the underlined sentence.” Responses were entered in the same way as for the no-context condition.

Participants were randomly assigned to one of three between-subjects conditions for the metaphor task (see Table 1; complete context materials are provided in Supplemental Online Materials, Appendix II). In the no-context condition all metaphors were presented without any context. In the metaphor-congruent condition, for each critical sentence a prior linguistic context established a theme supporting the intended metaphorical interpretation. These contexts were composed by the authors. In the literal-congruent condition, the linguistic context was designed to support a literal interpretation of the critical sentence, and thus to interfere with finding the metaphorical interpretation. Literal-congruent contexts provided encyclopedia-like information related to topics relevant to the target domain.

Results

Metaphor comprehension scores

Responses to the metaphor task were graded using a scoring key (see Table 1; also Supplemental Online Materials, Appendix I). The scoring key was constructed by two raters (D.S. and N.I.) based on a sample of 127 initial participant responses. A response was awarded 0, 1, or 2 points. A score of 2 (completely correct) was given if the paraphrase captured the metaphorical meaning at a level of abstraction beyond the source domain (i.e., a paraphrase that did not repeat the metaphorical formulation). A score of 1 (partially correct) was given if the paraphrase described the metaphorical meaning but maintained explicit links to the source domain. A score of 0 (incorrect) was given if the paraphrase was strongly linked to the source domain, merely ascribed simple physical features to the target, or rephrased the metaphor using simpler terms with no further insights (i.e., was literal in nature, constituting 32.1% of paraphrases receiving scores of 0), was a complete misinterpretation or nonsensical (62.6% of 0 scores), or if no response was entered (5.3% of 0 scores). The data for all three conditions were collected simultaneously; all raters graded responses as
a single undifferentiated batch, and were thus blind to condition, as well as to the results of the other tasks. Examples of different types of errors are provided in Supplemental Online Materials, Appendix III. Interrer agreement for the sample of responses scored by two raters was 95%. After the scoring key had been established, all other responses were scored by D.S., and all scoring was blind to condition. We calculated the mean score per item, thus placing the means on the 0–2 range used in scoring each individual item. These mean comprehension scores were 1.03 (SD = .21, skewness = .30, kurtosis = .59) for the no-context baseline condition, 1.18 (SD = .20, skewness = .29, kurtosis = .58) for the metaphor-congruent context, and 0.86 (SD = .26, skewness = .29, kurtosis = .57) for the literal-congruent context. Shapiro-Wilks tests of normality for comprehension scores within each condition revealed no violation of normality for any condition (no-context baseline: W(65) = .98, p = .323; metaphor-congruent context: W(69) = .97, p = .109; literal-congruent context: W(67) = .98, p = .227).

**Individual difference measures**

Table 2 summarizes the descriptive statistics for the three measures we obtained (RPM, SST, and WAIS-III Vocabulary), and Table 3 shows the intercorrelations among them. As would be expected given random assignment of participants, the pattern did not differ reliably across context conditions for any of these measures; hence Tables 2 and 3 present data from the entire set of 201 participants in the study. We used the cocor package in R to test the difference between these intercorrelations (Diederhofen & Musch, 2015). The cocor package implements ten distinct statistical methods to test the difference between correlation coefficients; for our analyses, all methods produced convergent results. For convenience, we report the results implementing Dunn and Clark’s (1969) z-test. The correlation between performance on our two measures of crystallized verbal intelligence, the WAIS-III Vocabulary measure and the SST (r = .471), was reliably greater than that between the WAIS-III and the RPM, our measure of fluid intelligence (r = .285, z = 2.59, p = .01), but not reliably greater than that between the SST and the RPM (r = .367, z = 1.41, p = .16). This pattern (also observed by Stamenković et al., 2019) is consistent with the interpretation that the RPM measures fluid intelligence whereas both the SST and WAIS-III Vocabulary tests measure crystallized intelligence. These two forms of cognitive ability are related to one another but separable (Cattell, 1971).

**Influence of context and individual differences on metaphor comprehension**

The next step was to examine the patterns of intercorrelations linking individual differences to metaphor comprehension across the three context conditions. For these analyses we standardized scores on all three individual-difference tests, and used the mean of the SST and WAIS-III Vocabulary scores to form a composite measure of crystallized intelligence (following the procedure employed by

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Max</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPM</td>
<td>6.78</td>
<td>12</td>
<td>2.87</td>
<td>0–12</td>
</tr>
<tr>
<td>SST</td>
<td>28.32</td>
<td>37</td>
<td>4.00</td>
<td>15–37</td>
</tr>
<tr>
<td>WAIS-III Vocabulary</td>
<td>47.16</td>
<td>62</td>
<td>8.31</td>
<td>6–62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>WAIS-III Vocabulary</th>
<th>SST</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPM</td>
<td>.285</td>
<td>.367</td>
</tr>
<tr>
<td>SST</td>
<td>.471</td>
<td></td>
</tr>
</tbody>
</table>

All p’s < .001
Stamenković et al., 2019). This combined measure allows for a more comprehensive measure of crystallized verbal knowledge than what is individually captured by the SST and WAIS-III scores. The RPM was used as a measure of fluid intelligence.

We performed a multiple regression analysis to examine the influence of both context and individual differences in fluid and crystallized verbal intelligence on metaphor comprehension within a unified model:

\[ M = C_M + C_L + I_F + I_C + C_M * I_F + C_M * I_C + C_L * I_F + C_L * I_C. \]

Here, \( M \) represents metaphor comprehension score; \( C_M \) represents metaphor-congruent context, compared to the no-context baseline; \( C_L \) represents the literal-congruent context, compared to the no-context baseline; \( I_F \) represents fluid intelligence (standardized RPM scores); and \( I_C \) represents crystallized verbal intelligence (standardized mean of standardized SST and WAIS-III scores). Product terms in the equation represent interactions between factors (e.g., \( C_M * I_F \) represents an interaction between metaphor-congruent context, compared to the no-context baseline, and fluid intelligence). A Levene’s test on metaphor comprehension scores across conditions found no violation of homogeneity of variances, \( F(2,198) = 1.65, p = .194 \), and our unified model significantly accounted for metaphor comprehension, \( F(8, 192) = 26.12, p < .001 \), adjusted R-squared = .50.

Examining specific effects within the model, we first considered the effect of context on metaphor comprehension. The analysis revealed a main effect for metaphor-congruent context compared to the no-context baseline, \( t(192) = 5.91, p < .001, \beta = .73 \), as well as a main effect for literal-congruent context compared to the no-context baseline, \( t(192) = 5.46, p < .001, \beta = -.68 \). These results support the prediction that relative to the no-context baseline, metaphor comprehension was facilitated by a metaphor-congruent context but impaired by a literal-congruent context.

Next, we examined the effects of individual differences in fluid and crystallized verbal intelligence on metaphor comprehension scores. Figure 1 shows our individual difference measures plotted against metaphor comprehension scores for each context. There was a reliable effect of fluid intelligence on metaphor comprehension score in the no-context baseline, \( t(192) = 2.04, p = .043, \beta = .20 \). Further, there was no interaction between fluid intelligence and the metaphor-congruent context compared to the no-context baseline, \( t(192) = 1.82, p = .070, \beta = -.24 \), or for literal-congruent context compared to the no-context baseline, \( t(192) = 1.11, p = .269, \beta = -.15 \). These results indicate that fluid intelligence was a reliable predictor of metaphor comprehension scores, and that individual contexts did not reliably differ in the extent to which fluid intelligence predicted metaphor comprehension.

In addition, there was a reliable effect of crystallized verbal intelligence on metaphor comprehension in the no-context baseline condition, \( t(192) = 2.89, p = .004, \beta = .29 \). Further, there was a reliable interaction of crystallized intelligence with metaphor-congruent context compared to the no-context baseline, \( t(192) = 2.66, p = .008, \beta = .37 \). However, there was no reliable interaction with the literal-congruent context compared to the no-context baseline, \( t(192) = 1.20, p = .233, \beta = .16 \). These findings highlight the association between crystallized verbal intelligence and metaphor comprehension across contexts, and indicate that the metaphor-congruent context increased the contribution of crystallized verbal intelligence. This increased reliance on semantic knowledge suggests that the metaphor-congruent context encouraged and facilitated semantic integration as the dominant strategy for interpreting the critical sentence.

Although both individual difference measures reliably accounted for separable variance in metaphor comprehension scores across contexts, we went on to directly compare the effects of fluid intelligence and crystallized intelligence on metaphor comprehension. A Wald test revealed that crystallized verbal intelligence was a stronger predictor than fluid intelligence in the no-context baseline, \( F(2,192) = 10.93, p < .001 \). Further Wald tests revealed the same pattern of effects for the metaphor-congruent, \( F(5, 192) = 19.24, p < .001 \), and the literal-congruent contexts, \( F(5, 189) = 16.61, p < .001 \). These findings replicate those of Stamenković et al. (2019) for the no-context baseline, and extend them to metaphor-congruent and literal-congruent contexts.
Influence of context on response types for metaphor comprehension

The fact that a literal-congruent context led to lower metaphor comprehension scores, despite the absence of any time pressure, suggests that participants were unable to apply an effective strategy to overcome the literal or misleading orientation evoked by this type of context. To provide a more detailed picture of the influence of the alternative contexts, Figure 2 shows the proportion of responses for each context condition that received scores of 0 (incorrect), 1 (partially correct), and 2 (completely correct). We conducted three one-way ANOVAs across context conditions (performed separately for 0, 1 and 2 scores). In order to control FWER at .05 across these three separate ANOVA’s, we used a more stringent, adjusted alpha of .017 (.05/3 = .017). These tests yielded a significant main effect for the number of 0 scores, $F(2,198) = 31.27$, $p < .001$, $\eta^2 = 0.24$, as well as for the number of 2 scores, $F(2,198) = 22.36$, $p < .001$, $\eta^2 = 0.18$. The number of 1 scores also varied across conditions, although the effect size was smaller, $F(2,198) = 4.26$, $p = .015$, $\eta^2 = 0.04$. Relative to the no-context baseline, the metaphor-congruent and literal-congruent contexts generated mirror-image response patterns for 0 and 2 scores. Among the three conditions, the metaphor-congruent context yielded the lowest proportion of 0 scores and the highest proportion of 2 scores, whereas the literal-congruent context yielded the highest proportion of 0 scores and the lowest proportion of 2 scores.

A more detailed analysis was performed on the frequencies of the two major subtypes responses receiving a score of 0 on the metaphor comprehension task, literal responses and misinterpretations. A literal response interpreted the target metaphor as a literal sentence, where the words or phrases representing the target and/or source concepts retain their respective conventional meanings, as opposed to being combined to form an abstracted, metaphorical meaning. For example, for the target metaphor “Nerves after a quarrel are frozen leaves in winter,” where “nerves after a quarrel” represents the target concept and “frozen leaves in winter” represents the source concept, literal responses included, “Frozen leaves are still but very delicate,” “Nerves after fighting are cold,” and “Nerves are stuck.” In contrast, a misinterpretation of a target metaphor significantly diverged from both a metaphorical and a literal interpretation. For example, given the same target metaphor, misinterpretations included, “Nerves usually tend to calm down after a fight,” “All your nerves just stop working when encounter something bad,” and “Nerves after a quarrel are not easily changed.” Further examples are provided in Supplemental Online Materials, Appendix III.
Figure 2. Proportion of responses receiving each metaphor comprehension score (0, 1 or 2) for each of the three context conditions. Error bars indicate ± standard error of the mean.

Shapiro-Wilks tests of normality for literal responses within each condition revealed violations of normality for the no-context baseline, $W(65) = .84$, $p > .001$; the metaphor-congruent context, $W (69) = .73$, $p > .001$; and the literal-congruent context, $W(67) = .95$, $p = .005$. A Kruskal-Wallace test of context on the frequency of literal responses revealed a significant difference across the three conditions, $H(2) = 70.91$, $p < .001$. Post-hoc pairwise comparisons with a Bonferroni correction to control FWER at .05 revealed that participants produced more literal responses in the literal-congruent condition (mean rank = 144.24) than in either the no-context (mean rank = 96.88, Standardized $U = 4.81$, $p > .001$) or metaphor-congruent conditions (mean rank = 62.89, Standardized $U = 8.39$, $p > .001$), and produced more literal responses in the no-context condition than in the metaphor-congruent condition (Standardized $U = 3.48$, $p = .002$).

Shapiro-Wilks tests of normality for misinterpretations within each condition revealed violations of normality for the no-context baseline, $W(65) = .93$, $p > .001$; the metaphor-congruent context, $W (69) = .87$, $p > .001$; and the literal-congruent context, $W(67) = .96$, $p = .043$. A Kruskal-Wallace test of the effect of context on the frequency of misinterpretations revealed a significant difference across the three conditions, $H(2) = 18.73$, $p < .001$. Post-hoc pairwise comparisons with a Bonferroni correction to control FWER at .05 revealed that participants produced more misinterpretations in the literal-congruent condition (mean rank = 123.60) than in either the no-context (mean rank = 98.94, Standardized $U = 2.46$, $p = .042$) or the metaphor-congruent conditions (mean rank = 80.99, Standardized $U = 4.31$, $p > .001$). The number of misinterpretations did not differ reliably across the no-context condition and the metaphor-congruent condition (Standardized $U = 1.80$, $p = .214$).

Thus, the literal-congruent condition evoked a greater frequency of both literal interpretations of metaphors and complete misinterpretations of them. This overall pattern is consistent with the hypothesis that both types of contexts encouraged a semantic integration strategy. However, whereas a metaphor-congruent context favors semantic integration consistent with the metaphorical interpretation, a literal-congruent context impairs the integration process. Sometimes this interference results in a literal interpretation, but often the inherent conflict between literal and metaphorical interpretations results in a complete misinterpretation.

**Discussion**

**Summary and implications**

The present findings reveal that the impact of individual differences on metaphor comprehension is modulated by prior context. Overall, comprehension of literary metaphors is positively related to measures of both fluid and crystallized intelligence, confirming and extending the previous findings of
Stamenković et al. (2019). When a literary metaphor is preceded by a metaphor-congruent context, the predictive power of crystalized intelligence is enhanced. These findings support the hypothesis that a supportive linguistic context encourages use of semantic integration in interpreting metaphors.

The observed patterns of individual differences have implications for understanding the cognitive mechanisms that may underlie metaphor comprehension. For several decades, two general theoretical proposals – analogical reasoning and semantic integration – have dominated discussion. Analogical reasoning appears to place substantial demands on working memory and inhibitory control – i.e., domain-general executive functions closely linked to fluid intelligence. In contrast, semantic integration is viewed as more language-specific, and particularly dependent on semantic knowledge. If integration can be performed by operations on vectors of semantic features (e.g., Kintsch & Mangalath, 2011), without necessitating explicit manipulation of relations in working memory, then the burden on executive functions will be reduced relative to the requirements for analogical reasoning.

Thus, we would expect individual differences in fluid intelligence to have a strong association with metaphor comprehension when an analogy strategy plays a prominent role, whereas individual differences in crystalized intelligence are more likely to dominate when semantic integration is the primary strategy. From previous work using metaphors presented in isolation (Stamenković et al., 2019), it appears that crystalized intelligence predicts successful comprehension for both literary and nonliterary metaphors, whereas fluid intelligence is primarily predictive for more complex literary metaphors of the sort used in the present study. These findings suggest that the role of analogical reasoning may be limited to comprehension of more complex metaphors, whereas semantic integration plays a more general role for all metaphor types. The present findings demonstrate that for literary metaphors, the impact of variations in crystalized intelligence is enhanced when the metaphor is preceded by a supportive linguistic context. Taken as a whole, the evidence from individual-difference data implies that semantic integration may be involved in grasping all types of metaphors, whether encountered in isolation or in a meaningful context, but is especially important when a supportive linguistic context is provided.

An intriguing aspect of the present findings is that even an unfavorable context – one that is congruent with a literal rather than metaphorical interpretation – apparently involves a comprehension strategy based at least in part on semantic integration. The general tendency to rely on semantic integration is consistent with the view that integration processes are fundamental to comprehension of meaningful text, whether its content is literal, metaphorical, or some mix of both (Kintsch, 1998; Kintsch & Mangalath, 2011). In other words, people naturally apply semantic integration at all times when comprehending text, supplementing this strategy with more explicit reasoning processes mainly under special circumstances (e.g., upon encountering a complex metaphor devoid of context). Our findings are consistent with those of a recent neuroimaging study that did not find differences between the neural processing of metaphors and matched literal sentences when both were embedded in an appropriate narrative context (Hartung et al., 2020).

**Directions for future research**

The current study has several limitations. One important issue stems from the nature of the metaphor comprehension task used in the present study. To provide correct answers on this type of open-ended metaphor comprehension task, it is necessary to both comprehend the metaphor and express its meaning verbally. It is possible that those with higher crystalized intelligence scores are better able to construct paraphrases of metaphor meaning. It would be useful to perform follow-up studies using other measures of comprehension (e.g., multiple-choice format), so as to more clearly separate comprehension ability from verbal expression.

It would also be valuable to compare the impact of different types of contexts, especially for literary metaphors. A recent study of the time course and neurophysiological underpinnings of metaphors in literary contexts suggests that such contexts enable the manipulation of multiple meanings in working memory (Bambini, Canal, Resta, & Grimaldi, 2019). Especially for novel metaphors, the simultaneous evocation of multiple meanings creates what have been termed plurisigns (Wheelwright, 1968), which
yield poetic effects (Holyoak, 2019). A literary context may also serve to establish a literary mode that facilitates the recognition of metaphors (Steen, 1994).

The individual-differences approach has proved useful in identifying mechanisms involved in understanding literary and nonliterary metaphors; however, many important questions remain open. Literary metaphors, which reflect the most creative aspects of language, can be further differentiated based on a variety of cognitive and emotional characteristics (Holyoak, 2019; Jacobs, 2015; Jacobs & Kinder, 2018). For all types of metaphors, variations in such factors as aptness, familiarity, syntactic complexity, and source-target similarity appear to influence comprehension (see Holyoak & Stamenković, 2018; Roncero & de Almeida, 2015). Future research should attempt to tease apart the contributions of different dimensions of variation among metaphors. Such differences include variations in the syntactic forms of metaphors (e.g., Cardillo et al., 2012; Li, Zhu, & Wang, 2013; Rai & Chakraverty, 2020), as well as varying levels of abstraction (Banaruee, Khoshimia, Zare-Behtash, & Yarahmadzai, 2019).

Another valuable extension of the individual-differences approach would be to examine metaphor comprehension and production tasks with the same participants. Given the strong association that exists between fluid intelligence and working memory capacity (e.g., Kane, Hambrick, & Conway, 2005), and the role that executive functions play in metaphor comprehension (e.g., Chiappe & Chiappe, 2007; Pierce & Chiappe, 2008), future work on individual differences should include tests of working memory. The individual-differences approach can also be combined with measures of creative potential (e.g., Faust & Kenett, 2014; Kenett, Gold, & Faust, 2018) to provide further insights into literary metaphor comprehension. Other individual differences measures, such as measures of personality traits, preference for imagery, and emotional understanding (see Fetterman, Bair, Werth, Landkammer, & Robinson, 2016) could provide more comprehensive understanding of the range of factors that impact metaphor comprehension and production. It would also be extremely interesting to compare the performance of different age groups. Although fluid intelligence is reduced with normal aging, the type of semantic knowledge underlying crystalized intelligence is spared or even enhanced (Salthouse, 2004). Thus older adults may maintain their abilities in metaphor processing, even though analogical reasoning with formal materials is impaired (Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004).

In addition, future research should consider metaphorical processing in languages other than English. Attention to cross-linguistic differences may prove particularly revealing, given the emerging picture of metaphor comprehension as heavily dependent on language-specific processes encompassed by semantic integration. Cognitively relevant differences across particular languages in their syntax or lexicon may offer useful avenues to illuminate how human reasoners combine lexical units to produce metaphorical speech or text, and to extract contextualized meaning from such speech or text. Another direction of future research within the individual-differences approach could relate it to pictorial and multimodal metaphors.

A deeper understanding of the roles played by the many factors that distinguish metaphors within a language as well as across languages will require the development of new computational models of metaphor processing (for a recent survey of computational approaches to metaphor comprehension, see Rai & Chakraverty, 2020). In this regard, recent work in artificial intelligence may prove to be relevant. In particular, machine-learning algorithms applied to large text corpora are able to create feature vectors representing the meanings of individual words (e.g., Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), as well as of larger linguistic units such as sentences (Devlin, Chang, Lee, & Toutanova, 2018). These tools, which generate feature-based semantic representations that can be merged to create contextualized meanings, may prove to be applicable to some aspects of semantic integration involved in metaphor comprehension.

**Acknowledgments**

Preparation of this paper was supported by the Faculty of Philosophy (University of Niš) internal grant No 360/1-16-1-01 to Dušan Stamenković and by NSF Grant BCS-1827374 to Keith Holyoak. We thank Aliza Ayaz, Diego Fernandez-Salvador, and Anshu Patel for assistance in data collection.
Disclosure statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

This work was supported by the NSF [BCS-1827374]; Faculty of Philosophy (University of Niš) internal grant [360/1-16-1-01].

ORCID

Dušan Stamenković http://orcid.org/0000-0002-0121-4591
Nicholas Ichien http://orcid.org/0000-0002-0928-0809
Keith J. Holyoak http://orcid.org/0000-0001-8010-6267

References


