Journal of Memory and Language 105 (2019) 108–118

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Metaphor comprehension: An individual-differences approach☆

Dušan Stamenković[a,*], Nicholas Ichien[b], Keith J. Holyoak[b]

[a] Faculty of Philosophy & Center for Cognitive Sciences, University of Niš, Serbia
[b] Department of Psychology, University of California, Los Angeles, United States

ARTICLE INFO

Keywords:
Metaphor
Analogy
Conceptual combination
Individual differences
Fluid intelligence
Crystalized intelligence

ABSTRACT

The nature of the mental processes involved in metaphor comprehension has been the focus of debate, with controversy focusing on the relative role of general analogical reasoning versus language-specific conceptual combination. In the present set of studies, we take an individual-differences approach to examine the comprehension of a variety of metaphors, some taken from literary sources, using several types of comprehension tests. In a series of metaphor-comprehension studies with college students, we measured both fluid intelligence (using the nonverbal Raven's Progressive Matrices test) and crystalized verbal intelligence (using a new Semantic Similarities Test as well as the Vocabulary subscale of the Wechsler Adult Intelligence Scale). Previous work has shown that measures of fluid intelligence are closely linked to individual differences in analogical reasoning, whereas measures of crystalized verbal intelligence are linked to language-specific abilities. We found that each measure had a dissociable predictive relationship to metaphor comprehension. The pattern of individual differences indicated that crystalized intelligence influences metaphor comprehension across a broad range of metaphor types, whereas individual differences in fluid intelligence mainly impact comprehension of more cognitively complex metaphors, such as those that arise in literary sources.

Introduction

Metaphor is the use of language to describe one thing in terms of something else that is conceptually very different, as in the poet Theodore Roethke's lament, "my memory, my prison." Metaphor and related cognitive processes have been linked to creative thinking, not only in poetry (Holyoak, 1982, 2019), but also in many scientific fields (e.g., Dunbar & Klahr, 2012). In artificial intelligence, the goal of automatically detecting and comprehending metaphors encountered in text corpora represents a current frontier (e.g., Gagliano, Paul, Booten, & Hearst, 2016). Given its evident importance in human thinking and language, an important goal for cognitive science is to understand how people grasp metaphors.

Potential mechanisms underlying metaphor comprehension

Psychologists, linguists, and philosophers have advanced many alternative theories of metaphor comprehension (e.g., Bowdle & Gentner, 2005; Gentner & Clement, 1988; Glucksberg & Keysar, 1990; Lakoff & Johnson, 1980; Ortony, 1979). Two general accounts have been especially influential. One proposal is that metaphor comprehension requires analogical reasoning to relate the target[1] to the source. The idea that metaphor is based on analogy originated with Aristotle and was advanced in modern times by Black (1962). In psychology, this hypothesis was developed further by Tourangeau and Sternberg (1981, 1982), Trick and Katz (1986), and Gentner and Clement (1988). Although computational models of analogical reasoning differ in important ways (e.g., Falkenhainer, Forbus & Gentner, 1989; Hummel & Holyoak, 1997, 2003), the general view is that the source and target are each represented as complex propositional structures, and that a process of *mapping* identifies systematic correspondences between elements of the two structures. Thus to comprehend Roethke's metaphor, a person might represent *memory* in terms of propositions such as "mental space in which information is stored for later retrieval", and *prison* as "secure building in which prisoners are confined against their will for long periods of time." A mapping process might then link "mental space" to "secure building" and "information" to "prisoner", etc. Because the mapping process depends on active manipulation of complex multi-component structures, and consideration of multiple alternative hypotheses, this sort of explicit analogical reasoning places a high load on working memory and executive functions such as attentional control (e.g., Cho, Holyoak, & Cannon, 2007; Waltz, Lau, Grewal, & Holyoak, 2000).

☆ The research data for this paper is stored in Mendeley Data: https://data.mendeley.com/datasets/4m7d49hdxs/1.
* Corresponding author at: Faculty of Philosophy & Center for Cognitive Sciences, University of Niš, Kabinet 418, Ćirila i Metodija 2, 18000 Niš, Serbia.
E-mail addresses: dusan.stamenkovic@filfak.ni.ac.rs (D. Stamenković), ichien@ucla.edu (N. Ichien), holyoak@lifesci.ucla.edu (K.J. Holyoak).
[1] The terms *topic* and *vehicle* are sometimes used for what we term the *target* and *source*. The latter terms are commonly used in discussions of analogical reasoning.

https://doi.org/10.1016/j.jml.2018.12.003
Received 17 February 2018; Received in revised form 4 December 2018

An alternative account, proposed by Glucksberg and Keysar (1990), and advanced by several other authors (e.g., Gernsbacher, Keysar, Robertson & Werner, 2001; McGlone & Manfredi, 2001; Jones & Estes, 2005), claims that metaphors (at least those in the nominal format, < *noun 1* > *is* < *noun 2* >) are interpreted as categorization statements. On this view, when Roethke claims that his memory is a prison, the target (*memory*) is stated to be a member of a category specified by the source (*prison*), where the latter takes on an abstract meaning like "location of extended confinement," rather than its more specific meaning of a building that houses prisoners.

The distinction between the analogy and categorization views can be interpreted more broadly as a distinction between analogy and *conceptual combination* (Holyoak & Stamenković, 2018). In general, sentence meanings (whether literal or metaphorical) are understood as systematic combinations of the meanings of constituent words. A great deal of evidence indicates that when people try to make sense of novel noun-noun combinations (e.g., "robin hawk"), they often interpret the modifier noun much like an adjective, extracting some salient property from it, which is than applied to the head noun (e.g., Wisniewski, 1997; Wisniewski & Love, 1998). Thus a robin hawk might be a kind of hawk with a red breast similar to that of a robin. Estes and Glucksberg (2000; also Gagné, 2002) argued that the categorization view of metaphor can be interpreted as a type of conceptual combination, in which the head concept provides relevant dimensions and the modifier concept provides candidate features for attribution.

Analogy and conceptual combination both rely on decomposing the source and target into elements, which are then compared and somehow integrated so as to create coherence. However, it has been argued that conceptual combination cannot be reduced to analogy (Keane & Costello, 2001). Whereas analogical reasoning is typically considered to be a domain-general process that operates on complex propositional structures held in working memory (Holyoak, 2012), conceptual combination is viewed as a simpler process based on spreading activation that operates at the level of lexical concepts (Kintsch, 2000, 2001; Kintsch & Bowles, 2002).

Despite decades of research addressing the question of whether metaphor comprehension depends on analogy, conceptual combination, or some mix of both (e.g., Bowdle & Gentner, 2005), no firm answer has emerged (for recent reviews see Kertész, Rákosi & Csatár, 2012; Patterson, 2016; Holyoak & Stamenković, 2018). Psychological studies have largely focused on simple nominal metaphors (e.g., *The lawyer is a shark*). In general, such metaphors appear to be processed relatively easily, whereas analogical reasoning (at least when performed explicitly) tends to place a heavy burden on working memory and executive processes (Holyoak, 2012). Kintsch and Bowles (2002) argued that for cases in which metaphor comprehension appears to be relatively easy, complex analogical reasoning is not a viable mechanism. At the same time, even proponents of the conceptual combination view have cautioned that this process is not sufficient for comprehension of metaphors of high cognitive complexity: cases in which the source and target are more semantically distant, the syntax is relatively complex, and the interpretation depends on finding correspondences between multiple elements of the two concepts (Glucksberg & Haught, 2006; Kintsch, 2000).

Neural evidence may be useful in discriminating between analogical reasoning and conceptual combination. Neuropsychological (e.g., Waltz et al., 1999; Morrison et al., 2004) and neuroimaging studies (e.g., Bunge, Helskog, & Wendelken, 2009; Cho et al., 2010) have established that complex analogical reasoning involves broad regions of the frontal and parietal cortices that form a frontoparietal network (Duncan, 2010). In particular, numerous studies (e.g., Bunge et al., 2009) have shown that complex analogical reasoning (including reasoning about verbal analogies that cross semantic domains, as metaphor does; Green et al., 2010, 2012) is almost invariably accompanied by activation of the left rostrolateral prefrontal cortex (RLPFC). Meta-analyses of neuroimaging studies reported by Vartanian (2012) and by Hobeika et al. (2016) support this conclusion. In contrast, conceptual combination (as applied to literal word meanings) primarily activates the left anterior temporal lobe (Baron & Osherson, 2011; Baron, Thompson-Schill, Weber, & Osherson, 2010), a region viewed as a "semantic hub" (Hoffman, McClelland, & Lambon Ralph, 2018).

Most neuroimaging evidence indicates that simple metaphors can generally be comprehended without involvement of the brain area most closely linked to complex analogical reasoning, rostrolateral prefrontal cortex (see meta-analyses by Bohrn, Altmann & Jacobs, 2012; Rapp, Mutschler, & Erb, 2012; Vartanian, 2012), even when the metaphor is novel (Cardillo et al., 2012). However, at least one study found evidence that neural correlates of metaphor comprehension largely overlap with those for analogical reasoning, including (in limited conditions) the rostrolateral prefrontal cortex (Prat, Mason & Just, 2012).

Individual-differences approach

Rather than assuming metaphor comprehension to be a unitary process that is constant across individuals, in the present study we adopted an individual-differences approach. Classical theories of intelligence (Cattell, 1971) distinguish *fluid* and *crystalized* intelligence as separable factors (although they tend to be moderately correlated with one another). Fluid intelligence involves reasoning (often nonverbal) about novel problems detached from prior knowledge, and crystalized intelligence involves reasoning (typically verbal) that draws upon prior knowledge. Metaphor comprehension seems likely to tap both of these basic forms of intelligence. Fluid intelligence is closely linked to analogical reasoning (Holyoak, 2012), whereas verbal crystalized intelligence is likely to impact conceptual combination (which is postulated to depend on lexical semantics). In order to measure crystalized verbal intelligence as it may relate to metaphor comprehension, we developed a new *Semantic Similarities Test* (SST).

A relatively small number of previous studies have investigated individual differences in cognitive factors that might impact processing of metaphors. Olkoniemi, Ranta, and Kaakinen (2016) assessed individual differences in the processing of metaphor and sarcasm using eye-tracking methods. These investigators found that individual differences in working-memory capacity and in cognitive style (Need for Cognition scale) were related to metaphor processing. Different eye-tracking patterns were observed for reading metaphors versus sarcasm, suggesting that these two forms of figurative language are processed in different ways. However, no assessment of crystalized intelligence was administered.

Trick and Katz (1986) found positive correlations between people's scores on a test of analogical reasoning and ratings of the comprehensibility of metaphors, especially when the source and target were drawn from dissimilar categories. A measure of vocabulary knowledge (which would be expected to reflect crystalized intelligence) did not add any predictive power. Nippold and Sullivan (1987) reported that within a sample of children, perceptual analogical reasoning was related to verbal analogical reasoning, as well as to comprehension of proportional metaphors (albeit weakly). A measure of verbal analogical reasoning did not independently contribute to prediction of metaphor comprehension. Thus neither of these studies provided support for a role of crystalized verbal intelligence in metaphor comprehension.

Kazmerski, Blasko, and Dessalegn (2003) had their participants complete IQ and working-memory tests and then rate and interpret a set of metaphors. The IQ measure included both fluid and crystalized components. They found that low-IQ participants produced poorer-quality interpretations relative to high-IQ individuals. A vocabulary subtest predicted interpretation quality (in apparent contrast to the null finding reported by Trick & Katz, 1986). However, a measure of spatial working-memory did not correlate with verbal IQ and did not predict quality of metaphor interpretations (a finding apparently contrary to that reported by Nippold & Sullivan, 1987). Thus although overall IQ predicted quality of metaphor interpretations, Kazmerski et al.'s findings did not clearly distinguish the impact of fluid and crystalized intelligence.

In a study by Chiappe and Chiappe (2007), individuals who scored high on a working-memory test generated higher-quality interpretations of metaphors more quickly. In addition (i.e., statistically separable from the impact of the working-memory measure), measures of inhibitory control (based on Stroop interference and intrusion errors on a memory test) also 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 found that measures of vocabulary knowledge and exposure to print (linked to crystalized intelligence) also predicted metaphor quality. Indeed, the measures of crystalized intelligence yielded somewhat higher correlations with metaphor than did the measures of working memory. Thus although findings have been mixed, at least the study by Chiappe and Chiappe (2007) suggests that both fluid and crystalized intelligence have an impact on metaphor processing.

Overview of present study

The present study sought additional evidence of potential individual differences in metaphor comprehension. In three studies, we assessed college students' ability to comprehend a variety of metaphors, relating their performance to measures of both fluid and crystalized intelligence. In all studies fluid intelligence was measured using a version of the Raven's Progressive Matrices test (RPM; Raven, 1938), which is highly correlated with performance in tests of analogical reasoning (Snow, Kyllonen, & Marshalek, 1984). In order to measure crystalized verbal intelligence as it may relate to metaphor comprehension, we developed a new *Semantic Similarities Test* (SST) (see Appendix), which was used in all studies. In Study 3 we also administered a standard test of verbal crystalized intelligence, the Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS-III).

Given the frequent claims that analogical reasoning (and hence fluid intelligence) is more likely to play a role in comprehending metaphors that are in some way conceptually complex (e.g., Kintsch, 2000; see Holyoak & Stamenković, 2018), we examined a range of metaphors selected to vary in complexity, or more generally in difficulty. We also examined a range of syntactic forms for single-sentence metaphors, including proportional metaphors based on nouns (e.g., *The eagle is a lion among birds)* and predicate metaphors (e.g., *The violent image rattled in her head*).

Most psychological studies of metaphor comprehension have focused on metaphors constructed by researchers, rather than those found in poetry and other literary writing (but see Tourangeau & Rips, 1991). In Studies 1–2 we compared comprehension of metaphors drawn from literary as well as nonliterary sources. The literary metaphors were selected from a large set 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 (i.e., the great majority of the stimulus sets used in psychological studies). Although the 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. These properties collectively suggest that good literary metaphors are high in cognitive complexity, which may be more likely to elicit analogical reasoning (and hence place greater demands on fluid intelligence). To provide a different manipulation of metaphor difficulty, Study 3 varied the familiarity of metaphors within a nonliterary set.

Study 1

Study 1 had two aims. First, we wished to compare the predictive power of the RPM (a measure of fluid intelligence) with that of a test designed to assess crystalized verbal intelligence. Second, we wished to examine comprehension of a wide range of metaphors, including a set derived from literary sources.

Method

Participants

A total of 76 UCLA undergraduates at the University of California, Los Angeles (UCLA) (female = 50, male = 25, undeclared = 1; mean age = 21.1) participated in the study for course credit. The great majority (91%) were native speakers of English, with a minority of bilinguals who spoke English fluently (self-assessed). The session generally took about 20–30 min to complete. Data from an additional five participants were dropped from analyses based on criteria indicative of carelessness or inattention on the verbal tasks: score of 12 or lower on the Semantic Similarities Test (max = 40), or 5 or lower (max = 20) on each set of metaphors, or extremely short overall response time (under 15 min for the entire set of tasks).

Design, materials, and procedure

Participants completed three tasks in a fixed order.

Raven's Progressive Matrices (RPM). A short version of the RPM (Arthur, Tubre, Paul, & Sanchez-Ku, 1999), adapted for computer administration using SuperLab software, was administered. The RPM is generally considered a central measure of fluid intelligence (Snow et al., 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) and spontaneous analogical transfer in a problem-solving task (Kubricht, Lu & Holyoak, 2017). To the best of our knowledge the RPM has never been used previously in conjunction with a test of metaphor comprehension. To the extent that analogical reasoning is required to comprehend metaphors, the RPM should be a robust predictor of individual differences in metaphor comprehension.

Semantic similarities test (SST). We created a new instrument to provide a rapid assessment of crystalized verbal intelligence with face validity as relevant to metaphor comprehension. The SST is designed to measure participants' ability to identify similarities between concepts expressed as single words, where the similarities varied in degree of generality and metaphoricity (see Appendix for complete test and answer key). 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. To create the test, we began by selecting five pairs with relatively specific similarities (e.g., *bird – airplane*, which both fly), five with more general similarities (e.g., *tavern – church*, which are both public buildings), and ten with more metaphorical similarities (e.g., *marriage – alloy*, which are both bonds between two elements). Across all pairs, we tried to make the types of similarity as diverse as possible. The items thus involve physical/perceptual similarities (e.g., orange-ball, peak-needle), functional similarities (e.g., sun-lightbulb, key-solution), structural similarities (e.g., corporation-tree, diamond-snowflake), taxonomic similarities (e.g., sword-pistol, tavern-church), and affective similarities (e.g., love-drug, memory-prison). These categories clearly overlap, and not intended as a strict taxonomy. Two of the investigators (DS and KH) discussed and altered items until a set of 20 items was generated that we agreed spanned the desired range of similarity types, and for which each word pair had at least one similarity that provides a compelling correct answer. The set

Table 1
Examples of literary and nonliterary metaphors in the metaphor comprehension task (Study 1).

Literary	Nonliterary
The expression *A body is a prison for the soul* means: (1) The soul can leave a body after death.* (2) The soul belongs to the body. (3) The soul is one with a body.	The expression *The nose is the antenna of scent* means: (1) The nose emits smells. (2) The nose resembles signals. (3) The nose can register smells.*
The expression *The tongue is a bayonet* means: (1) It can be difficult to talk. (2) The tongue is pointed. (3) Words can be hurtful.*	The expression *Invention is the child of an inventor* means: (1) Inventors are like children. (2) Inventors are the creators of inventions.* (3) Inventors neglect inventions.
The expression *The mind is a mountainous landscape* means: (1) It is difficult to understand how the mind operates.* (2) The mind allows us to imagine landscapes. (3) It is easy to understand how the mind operates.	The expression *Purgatory is the lobby of heaven* means: (1) One needs to go through heaven to get to purgatory. (2) One needs to go through purgatory to get to heaven.* (3) One needs to go through purgatory to get to the lobby.

* Indicates the response scored as correct.

Table 2
Descriptive statistics for each test (Study 1).

Test	Mean	Max	SD	Range
RPM	6.64	12	2.90	0–12
SST	29.03	40	4.05	19–37
Literary metaphors	15.90	20	2.91	6–20
Nonliterary metaphors	18.49	20	2.34	7–20

Table 3
Correlations and partial correlations of individual-difference measures with metaphor comprehension (Study 1).

	Literary		Nonliterary	
	Correlation	Partial correlation	Correlation	Partial correlation
RPM	.43***	.34**	.37**	.26*
SST	.43***	.34**	.49***	.42***

*** p < .001.
** p < .01.
* p < .05.

was piloted at UCLA and the University of Niš before it reached its final form.

For each pair, participants answered the question, "How are the two concepts in each pair similar to one another?" The instructions included a single example (*chair – sofa*), for which the answer provided was "both are types of furniture". An answer key was compiled based on pilot testing. Along with the answers we had anticipated, we broadened the key by classifying the plausible responses provided by those participating in the pilot stage as fully acceptable or partly acceptable (2 points and 1 point, respectively, with 0 points for incorrect responses). Thus the maximum possible test score is 40 points. Based on preliminary data, the items were ordered from easy-to-hard. The resulting ordering (see Appendix) generally places items with more metaphorical similarities towards the end, but with some overlap. Empirically, the easiest item of a metaphorical nature was ranked as number 8 (*love – drug*), and the most difficult non-metaphorical pair was ranked as number 17 (*diamond – snowflake*). The internal consistency of the scale was acceptable given the diversity of item similarities and lack of repeated items (Cronbach's alpha = 0.61, based on data from 280 participants; comparable to alpha = .69 for the short-form RPM based on the same participants).

Whereas the RPM is a formal and nonverbal test in which semantic knowledge plays virtually no role, the SST is a verbal test in which semantic knowledge of word meanings is critical. The RPM and SST thus complement each other as relatively pure measures of fluid and crystalized verbal intelligence, respectively. We would, however, expect scores on the two tests to be correlated, as both should load on the *g* factor (general intelligence; see Ackerman et al., 2005; Spearman, 1927).

Metaphor comprehension. The final task in this study consisted of 40 metaphor comprehension items, 20 from literary sources and 20 nonliterary. The literary metaphorical statements were selected from a list of literary metaphors drawn from poetry anthologies by Katz et al. (1988). The metaphors we chose were rated high on a goodness scale in the Katz et al. study (e.g., *The tongue is a bayonet*). Their syntactic forms included nominal (*X is Y*), nominal with an adjective modifier, and nominal with a prepositional phrase. (For detailed analyses of the properties of this metaphor set, see Jacobs & Kinder, 2017, 2018).

The nonliterary metaphors included 20 items, some of them adapted from word pairs generated by Green et al. (2010, 2012) to make proportional verbal analogy problems in the form *A:B*:: *C:D* (e.g., *roof:-house*:: *hat:man*). By dropping the D term, we converted some of these items into proportional metaphors in the form *A is the C of B* (e.g., *A roof is the hat of a house*). We augmented the set with similar items that we created following the same pattern. The literary and nonliterary items were intermixed and presented in a randomized order.

Comprehension was assessed by a task requiring selection of the best interpretation. For each metaphorical statement, three potential interpretations were provided, and the participants were asked to select the correct one. In this and subsequent studies, the multiple-choice items (including those labelled as correct options) were selected after discussion among authors and with the help of several independent raters. As literary metaphors may have several interpretations, the responses we identified as correct were options agreed upon as being one plausible interpretation. In selecting the remaining three response options (which were either literal or implausible), we ensured that the agreed interpretation would be the best among the alternatives offered. Examples of the interpretation task, for both literary and nonliterary metaphors, are shown in Table 1.

The stimuli for all three tasks were presented on a computer screen and participant responses were recorded. Instructions for each task were given immediately preceding that task. There was no time limit on any task, but participants were instructed to complete each task as quickly as possible.

Results and discussion

Performance on each task is summarized in Table 2. Correlation and regression analyses were performed to assess the interrelationships among the RPM, SST and metaphor comprehension. RPM and SST scores were moderately correlated with each other (r (76) = .31, p = .006). As summarized in Table 3, each individual-difference measure was correlated with accuracy on the comprehension test for both literary and nonliterary metaphors (individual correlations ranging from .37 to .49, p < .01 in all cases). Multiple regression

Table 4
Examples of literary and nonliterary metaphors in completion and interpretation tasks (Study 2).

Literary metaphor comprehension		Nonliterary metaphor comprehension	
Completion	Interpretation	Completion	Interpretation
Water is the blood of soft ———. (1) dreams (2) snows* (3) air	The expression *Water is the blood of soft snows* means: (1) Water brings coldness. (2) Water originates from soft snows.* (3) Soft snows are thicker than water.	An election is the ——— of votes. (1) cultivation (2) sowing (3) harvest*	The expression *An election is the harvest of votes* means: (1) Candidates collect signatures in an election. (2) Elections are scheduled at the same time as harvests. (3) Candidates collect votes in an election.*
——— is a leaf in the gardens of God. (1) Goddess (2) Man* (3) Mother	The expression *Man is a leaf in the gardens of God* means: (1) God cherishes human beings.* (2) God waters the soil. (3) Human beings love God.	A tire is the ——— of a car. (1) shoe* (2) ankle (3) elbow	The expression *A tire is the shoe of a car* means: (1) Tires and shoes have the same patterns. (2) Tires are made in the same way as shoes. (3) Tires help cars move on the ground.*
The ——— is a rope that binds heaven and earth. (1) mind (2) body (3) soul*	The expression *The soul is a rope that binds heaven and earth* means: (1) The soul contains both heaven and earth. (2) The soul is what makes heaven look like earth. (3) The soul allows one to travel from earth to heaven.*	——— is the morning of life. (1) Old age (2) Adulthood (3) Childhood*	The expression *Childhood is the morning of life* means: (1) Childhood is initiated before life. (2) Childhood comes at the same time as life. (3) Childhood comes early in life.*

* indicates the response scored as correct.

analyses revealed that for both types of metaphors, RPM and SST scores each predicted separable variance in comprehension accuracy, with partial correlations ranging from .26 (RPM for nonliterary metaphors, $p < .05$) to .42 (SST with nonliterary metaphors, $p < .001$). This pattern suggests that while both measures have an impact on metaphor comprehension, RPM (fluid intelligence) may be somewhat less important than SST (crystalized intelligence) for the simpler, nonliterary metaphors.[2]

Study 2A

Study 2 was designed to extend the findings of Study 1 by using multiple types of tasks to assess metaphor comprehension. In addition to the interpretation task used in Study 1 (select the best interpretation from a set of options), we also used a completion task (select the best word to complete a metaphor from a set of options). To avoid repeating items with different tasks, the 2 × 2 design (literary/nonliterary metaphors × interpretation/completion task) was decomposed into two pairs of conditions, which were run and analyzed separately. Table 4 shows examples of each type of metaphor with each comprehension task. The multiple-choice items were selected based on a discussion among authors and with the help of independent raters. Study 2A examined literary metaphors with the completion task and nonliterary metaphors with the comprehension task; Study 2B examined literary metaphors with the interpretation task and nonliterary metaphors with the completion task. We will introduce all four conditions as we describe Study 2A.

Method

Participants

A total of 101 undergraduate UCLA students (female = 77, male = 23, undeclared = 1; mean age = 20.1) participated in the study for course credit. The great majority (92%) were native speakers of English, with a minority of bilinguals who spoke English fluently (self-assessed). The session generally took about 20–30 min to complete. Data from an additional 11 participants were dropped from analyses based on criteria indicative of carelessness or inattention on the verbal tasks: score of 12 or lower on the SST (max = 40), or 4 or lower (max = 15) on each set of metaphors, or extremely short overall response time (under 15 min for the entire set of tasks).

Design, materials, and procedure

As in Study 1, all participants completed the RPM, SST, and metaphor comprehension tasks, in that order. The metaphors used in Studies 2A and 2B (15 literary and 15 nonliterary) were generally selected from among those used in Study 1, with a few revisions. We also revised some of the options, aiming to make them more challenging. Table 4 presents some examples.

In Study 2A, the comprehension task consisted of two conditions, administered in a fixed order. The first condition used literary metaphors with a completion task, in which each metaphor was presented with a blank (e.g., *Sunlight is a golden* ———.) for which a completion was to be chosen. Three options were presented underneath, one of which (scored as correct) was from the original metaphor (for this example, *dust*).

The second condition used 15 nonliterary metaphors with the task of choosing the best interpretation, as in Study 1. Within all metaphor-comprehension tasks, the items were displayed in a randomized order for each participant.

For all tasks, stimuli were displayed on a computer screen and participant responses were recorded. Participants received the instructions for each task separately, just before the relevant task. There was no time limit for any task, but participants were instructed to complete the task as quickly as possible.

Results and discussion

Performance on each task is summarized in Table 5A. As in Study 1, correlation and regression analyses were performed to assess the interrelationships among the RPM, SST and metaphor comprehension. RPM and SST scores were again reliably correlated with each other ($r(101) = .32, p = .001$). As summarized in Table 6A, both individual-difference measures were correlated with accuracy on the completion task with literary metaphors. Partial correlations revealed that each of the two individual-difference measures contributed separately to predicting performance for this condition. However, for the simpler nonliterary metaphors used in the interpretation task, only SST scores were

[2] Because scores on the RPM and SST were skewed, we also ran correlational analyses after performing a log transform on the two predictive tests. The basic pattern of correlations was unchanged by this transformation.

Table 5
Descriptive statistics for each test (Study 2).

Test	Mean	Max	SD	Range
A: Study 2A				
RPM	7.11	12	2.67	2–12
SST	30.02	40	3.83	15–37
Literary metaphors (completion)	10.10	15	1.88	5–15
Nonliterary metaphors (interpretation)	14.09	15	1.28	7–15
B: Study 2B				
RPM	6.55	12	2.27	1–11
SST	28.99	40	3.91	14–37
Literary metaphors (interpretation)	11.50	15	1.82	7–15
Nonliterary metaphors (completion)	12.02	15	1.66	7–15

Table 6
Correlations and partial correlations of individual-difference measures with metaphor comprehension (Study 2).

	Literary		Nonliterary	
	Correlation	Partial correlation	Correlation	Partial correlation
A: Pattern in Study 2A				
RPM	.38***	.30**	.19	.10
SST	.36***	.27**	.31**	.26**
B: Pattern in Study 2B				
RPM	.34***	.27**	.17	.08
SST	.30**	.21*	.32**	.28**

*** p < .001.
** p < .01.
* p < .05.

a reliable predictor of performance (based on both raw and partial correlations). The partial correlation obtained for the RPM was .10, $p = .34$.

Study 2B

Method

Study 2B was identical to Study 2A, except that the metaphor task consisted of the interpretation condition with nonliterary metaphors, followed by the completion condition with literary metaphors.

A total of 103 undergraduate UCLA students (female = 70, male = 33; mean age = 20.3) participated in the study for course credit. The great majority (94%) were native speakers of English, with a minority of bilinguals who spoke English fluently (self-assessed). The session generally took about 20–30 min to complete. Data from an additional nine participants were dropped from analyses based on the same criteria as in Study 2A.

Results and discussion

Performance on each task is summarized in Table 5B. Correlation and regression analyses were again performed to assess the interrelationships among the RPM, SST and metaphor comprehension. RPM and SST scores were again reliably correlated with each other (r (103) = .31, $p = .001$). As summarized in Table 6B, both individual-difference measures were correlated with accuracy on the interpretation task with literary metaphors. Partial correlations revealed that each of the two individual-difference measures contributed separately to predicting performance for this condition. For the simpler nonliterary metaphors used in the completion task, only SST yielded a significant raw correlation with metaphor performance, and also a reliable partial correlation. The partial correlation obtained for the RPM was .08, $p = .45$.

Study 3

The metaphors used in Studies 1–2 were based on nouns. Given that some of the items in the SST were derived from nominal metaphors, it could be argued that the SST proved to be a good predictor of metaphor comprehension in part because it was based on similar noun-to-noun comparisons.[3] In order to extend the generality of the findings from Studies 1–2, Study 3 examined comprehension of predicate metaphors in which the key word that takes on a metaphorical interpretation is a verb, rather than a noun. If the SST continues to predict comprehension even for predicate metaphors, this would imply that the test is measuring a more general ability than simply facility in comparing semantically dissimilar nouns.

A related concern is that because the SST is a novel test, we have yet to establish that it is a valid measure of crystalized intelligence (although the test has clear face validity as a measure of verbal semantic knowledge, being very similar in format to the Similarities subtest of the WAIS). To provide an explicit assessment of the SST as a measure of crystalized intelligence, in Study 3 we administered not only the RPM and SST, but also the Vocabulary subtest of the WAIS, which is a standard measure of verbal crystalized intelligence.

Finally, Study 3 used only nonliterary metaphors, but introduced a different factor expected to influence comprehension difficulty: familiarity. If any source of difficulty that impacts metaphor comprehension evokes analogical reasoning (and hence reliance on fluid intelligence), then the RPM would be expected to predict performance at least for less familiar metaphors. On the other hand, if the special sources of cognitive complexity associated with literary metaphors (Jacobs & Kinder, 2018) are key triggers for analogical reasoning, then the RPM may be a weak predictor of metaphor comprehension for the nonliterary metaphors used in Study 3, even for those that are relatively unfamiliar and hence difficult.

Method

Participants

A total of 85 undergraduate UCLA students (female = 58, male = 27, mean age = 19.1) participated in the study for course credit. The majority (79%) were native speakers of English, with a minority of bilinguals who spoke English fluently (self-assessed).[4] The session generally took about 20–35 min to complete. Data from an additional 6 participants were dropped from analyses based on criteria indicative of carelessness or inattention on the verbal tasks: score of 12 or lower on the SST (max = 40), or 3 or lower (max = 12) on each subset of metaphors, or extremely short overall response time (under 15 min for the entire set of tasks).

Design, materials, and procedure

As in the previous studies, participants first completed a short form

[3] To assess whether the predictive power of the SST is solely attributable to the items derived from metaphors, we divided the 20 test items into the subset of 10 derived from metaphors and the subset of 10 not derived from metaphors. Across all studies reported in this paper, we obtained comprehension scores for a total of 8 sets of metaphors. Each subset of the SST (considered separately) achieved a reliable correlation for 7 of these 8 metaphor conditions. The mean correlation over all metaphor conditions was .32 for the metaphor-derived subset of the SST, .26 for the non-metaphor-derived subset, and .35 for the full SST. Thus both subsets of items contributed about equally to the predictive power of the SST.

[4] Because the proportion of native speakers was somewhat lower in Study 3 than in the previous studies, we performed statistical tests to determine whether the native speakers and bilinguals differed on any of the language-related tests (SST, WAIS-III Vocabulary, and metaphor comprehension). Mean differences were very small in magnitude, and not statistically reliable for any of these measures.

of the RPM, followed by the SST. This was followed by the Vocabulary subtest of the third version of the Wechsler Adult Intelligence Scale (WAIS-III), a standardized measure of crystallized intelligence consisting of 33 vocabulary items. As in the case of the RPM, we adapted this test for computer administration using SuperLab software. Responses were assessed using the official scoring manual (Wechsler, 1997).

Predicate-metaphor comprehension. The metaphors used in Study 3 were all in a predicate format (e.g., *The violent image rattled in her head*), where the verb (*rattled*) was the focus of the metaphorical interpretation. All items were drawn from a nonliterary set created and normed by Cardillo et al. (2010). To systematically vary item difficulty, we selected metaphors that varied in familiarity according to the norms provided by Cardillo et al. To closely match metaphors in cognitive complexity across levels of familiarity, we chose all items from among a subset in which the verb was auditory in nature. To ensure that all metaphors were high in goodness, we restricted selection to those with a figurativeness rating of at least 4.5 (comparable to the goodness ratings used to select items from the norms of Katz et al., 1988, in Studies 1–2). Among the candidate items that satisfied the above requirements, we selected the 12 metaphors rated highest and the 12 rated lowest in familiarity, as normed by Cardillo et al.

Metaphor comprehension was assessed using a multiple-choice test that required selection of the best metaphorical interpretation from among three options. Options were developed by discussion among authors and with the help of independent raters. Examples of familiar and unfamiliar predicate metaphors and choice options are provided in Table 7.

For each task, stimuli were displayed on a computer screen and participant responses were recorded. Participants were given instructions for each task separately, immediately prior to the task itself. In the first three tasks (RPM, SST and WAIS-III Vocabulary) there was no time limit, but participants were instructed to complete the task as quickly as they could. For the metaphor comprehension task, the 24 metaphors were presented in a randomized order. Pilot work revealed that participants were close to ceiling in accuracy when allowed unrestricted time to select the correct option. Based on additional pilot testing, a deadline of 11 s to choose an option was imposed. Participants were informed of the time limit, and a warning beep sounded two seconds prior to the deadline. Response time to select an option was recorded.

Results and discussion

On the metaphor comprehension task, participants rarely failed to respond prior to the 11-s deadline (2% of trials for familiar metaphors, 4% for unfamiliar). A paired-samples *t*-test was conducted to compare comprehension scores (i.e., accuracy in selecting the metaphorical interpretation) for familiar versus unfamiliar predicate metaphors. Effect sizes for pairwise comparisons are reported using Cohen's *d*. Comprehension scores were reliably higher for the familiar ($M = 10.55$, $SD = 1.47$) than the unfamiliar metaphors ($M = 9.47$, $SD = 1.77$); $t(84) = 5.67$, $p < .001$, $d = .61$. Response times on correct trials also revealed an advantage for the familiar ($M = 5679$ ms, $SD = 978$ ms) relative to the unfamiliar metaphors ($M = 6778$ ms, $SD = 1044$ ms); $t(84) = 15.11$, $p < .001$, $d = 1.64$. These results confirm that lesser familiarity yielded greater difficulty in comprehending predicate metaphors.

Table 8
Descriptive statistics for each test (Study 3).

Test	Mean	Max	SD	Range
RPM	6.99	12	2.67	1–12
SST	30.26	38	3.36	23–38
WAIS-III Vocabulary	48.23	63	7.16	25–63
Familiar metaphors	10.55	12	1.48	4–12
Unfamiliar metaphors	9.47	12	1.77	4–12

Performance on each individual-difference measure, as well as comprehension scores for each metaphor type, are summarized in Table 8. As in Studies 1 and 2, correlation and regression analyses were performed to assess the interrelationships among the RPM, SST, WAIS-III Vocabulary subtest, and metaphor comprehension scores (see Table 9A). RPM and SST scores were again reliably correlated with each other, $r(85) = .39, p < .001$. As expected, SST scores were more strongly correlated with WAIS-III Vocabulary scores, $r(85) = .67, p < .001$, supporting the interpretation of the SST as a measure of verbal crystalized intelligence. RPM and WAIS-III Vocabulary scores were also reliably correlated, though less strongly, $r(85) = .42, p < .001$.

Given that both the SST and WAIS-III Vocabulary appear to measure verbal crystallized intelligence, we standardized scores on all three individual-difference tests (RPM, SST, and WAIS-III Vocabulary), and used the mean of the SST and WAIS-III Vocabulary scores to form a composite measure of crystalized intelligence. As summarized in Table 9B, this composite crystalized score yielded significant raw and partial correlations with comprehension scores for both familiar and unfamiliar metaphors. In contrast, the RPM was not reliably correlated with comprehension for either metaphor type. For familiar metaphors the partial correlation obtained for the RPM was .01, $p = .93$; for unfamiliar metaphors it was .03, $p = .78$. A parallel analysis using response times as the dependent measure yielded the same pattern. The failure of the RPM to predict comprehension scores or response times, even for unfamiliar metaphors (which were demonstrably more difficult), suggests that sheer difficulty is not sufficient to trigger increased reliance on fluid intelligence to comprehend metaphors.

Table 7
Examples of familiar and unfamiliar predicate metaphors and their interpretations used in Study 3.

Familiar predicate metaphors	Unfamiliar predicate metaphors
The expression *The violent image rattled in her head* means: (1) The violent image made a strong and persistent impression on her.* (2) In her opinion the violent image was a masterpiece. (3) She thought the violent image looked like a rattle being shaken.	The expression *The flowers purred in the sunlight* means: (1) The flowers seemed to be enjoying the sunlight.* (2) There was a cat sitting among the flowers. (3) Some flowers emit cat-like sounds in sunlight.
The expression *The dinner fizzled after the first course* means: (1) The first course was the worst dish of the dinner. (2) Starting from the second course, the dinner did not go well.* (3) After the first course the dinner ended abruptly.	The expression *The overhead bin snorted at the large suitcase* means: (1) The large suitcase fell out of the overhead bin. (2) The door of the overhead bin made a snorting sound. (3) The large suitcase would not fit easily into the overhead bin.*
The expression *His brain whirred in his skull* means: (1) He had to undergo brain surgery. (2) Something was bothering him.* (3) He was absent-minded.	The expression *The suede jacket yelped in the rain* means: (1) Whoever wore the suede jacket enjoyed the rain. (2) The rain was likely to damage the suede jacket.* (3) The suede jacket was made to be worn in the rain.

* indicates the response scored as correct.

Table 9
Correlations and partial correlations of individual differences measures with metaphor comprehension scores (Study 3).

A: Raw correlations of each individual-difference measure with metaphor comprehension score.

	Familiar	Unfamiliar
RPM	.16	.16
SST	.30**	.30**
WAIS-III Vocabulary	.31**	.26*

B: Correlations and partial correlations of the standardized RPM and the composite crystallized intelligence measure with metaphor comprehension scores.

	Familiar Scores		Unfamiliar Scores	
	Correlation	Partial Correlation	Correlation	Partial Correlation
RPM (standardized)	.16	.01	.16	.03
Composite Crystalized	.34**	.30**	.31**	.26*

** $p < .01$.
* $p < .05$.

General discussion

Summary

In three studies reported here we took an individual-differences approach to examine the cognitive factors that impact comprehension of metaphors. In particular, we sought to identify separable contributions of fluid intelligence (assessed by the Raven's Progressive Matrices) and verbal crystalized intelligence (assessed by a new Semantic Similarities Test, and in Study 3 by the WAIS-III Vocabulary subtest). Based on prior work on the cognitive and neural mechanisms underlying analogical reasoning and conceptual combination, the former process appears to be more closely linked to fluid intelligence and the latter to crystalized intelligence. Thus, the observed pattern of individual differences may shed light on the longstanding question of whether metaphor comprehension is more dependent on analogical reasoning or on conceptual combination, as well as whether the preferred strategy varies across different types of metaphors.

Studies 1 and 2 examined comprehension of metaphors presented in a proportional format, in which nouns undergo metaphorical meaning shifts. We compared metaphors derived from literary sources, taken from a set collected and normed by Katz et al. (1988), with metaphors in similar syntactic forms derived from nonliterary sources (i.e., constructed by metaphor researchers). Based on extensive quantitative analyses (Jacobs & Kinder, 2017, 2018), literary metaphors rated high in goodness (the type we selected for our studies) have a number of qualities, such as high surprisal and low source-target similarity, which contribute to their greater cognitive complexity. Both fluid intelligence (assessed by the RPM) and crystalized intelligence (assessed by the SST) yielded reliable and separable correlations with comprehension of literary metaphors. The SST also had a robust correlation with comprehension of nonliterary metaphors, for which the predictive power of the RPM was weak (Study 1) or unreliable (Study 2).

The weak correlation observed for the RPM in Study 1 for nonliterary metaphors may have been due to the fact that literary and nonliterary metaphors were randomly intermixed. An analogy strategy (dependent on fluid intelligence) elicited by the literary metaphors may have sometimes spilled over to trials with nonliterary metaphors. In Study 2 the literary and nonliterary metaphors were presented in separate blocks, making such spillover unlikely, and thus eliminating the correlation between the RPM and comprehension scores. In any case, the predictive power of the SST clearly exceeded that of the RPM for nonliterary metaphors.

In Study 3 we extended the range of stimuli by examining nonliterary predicate metaphors (where the focal word is a verb rather than a noun) that varied in familiarity, while holding cognitive complexity constant (by selecting familiar and unfamiliar metaphors using semantically-similar verbs related to audition). In addition to the RPM and SST, we administered the WAIS-III Vocabulary subtest, an established measure of crystalized intelligence. The latter two tests were highly correlated with one another, and hence their scores were combined to create a composite crystalized intelligence score. This measure reliably predicted metaphor comprehension (as well as response times) for both familiar and unfamiliar predicate metaphors. In contrast, the RPM was not a reliable predictor for either type.

Implications for models of metaphor comprehension

Given the strong association between the RPM and measures of analogical reasoning (Snow et al., 1984), the relative weakness of the connection between RPM scores and comprehension of simple metaphors casts further doubt on the hypothesis that complex analogical reasoning is necessary to understand such metaphors (Holyoak & Stamenković, 2018). Moreover, the finding in Study 3 that the RPM does not predict ease of comprehension even for unfamiliar (and more difficult) metaphors runs counter to the hypothesis that analogical reasoning is necessary to comprehend novel metaphors (Bowdle & Gentner, 2005). Similarly, in a neuroimaging study of the effect of novelty, Cardillo et al. (2012) were unable to detect greater involvement of brain areas related to analogy on the first encounter with a novel metaphor.

Our results are in accord with those of Chiappe and Chiappe (2007), who found evidence that both fluid and crystalized intelligence affect metaphor comprehension, with crystalized intelligence being the more potent factor (at least for simpler metaphors). In the present study, the factor that seemed to distinguish metaphors with a stronger link to analogical reasoning (i.e., a stronger correlation with RPM scores) was not novelty (as proposed by Bowdle & Gentner, 2005), but rather high cognitive complexity. Analogical reasoning may be required to comprehend more complex literary metaphors (consistent with the suggestions of Glucksberg & Haught, 2006, and Kintsch, 2000).

The joint influence of fluid and crystalized intelligence in comprehending metaphors, particularly those that are cognitively demanding, suggests that people may integrate multiple reasoning processes when dealing with metaphors. Processing strategies may change over the course of cognitive development and with the growth and refinement of lexical semantic representations. For example, Carriedo et al. (2016) found that the dependency of metaphor comprehension on executive functions declined between adolescents and young adults. Similarly, in a neuroimaging study, Prat et al. (2012) found that for adults, reading experience (likely to be linked to vocabulary growth and crystalized intelligence) was more strongly related to neural efficiency of metaphor comprehension than was working-memory capacity. To the extent that reading experience causes metaphor comprehension to become relatively more dependent on crystalized than on fluid intelligence, we would expect metaphor comprehension to be less susceptible to age-related declines in fluid intelligence. Indeed, whereas older adults exhibit deficits in solving formal analogy problems (Viskontas et al., 2004), at least some aspects of metaphor processing appear to be preserved in older individuals (Newsome & Glucksberg, 2002).

Directions for future research

Metaphor theorists have typically viewed analogy and conceptual combination as rival accounts. However, an intriguing possibility that warrants further investigation is that analogical reasoning and conceptual combination may operate together in processing more complex metaphors. Conceptual combination operates on featural representations of a pair of individual words, merging them in a constrained fashion to create a new representation (Kintsch, 2000, 2001). If analogical mapping is performed on a complex metaphor, the process will output paired source-target concepts (often corresponding to lexicalized word meanings). Each individual mapping will establish what has more generally been termed a *coupling* between words (Levin, 1962)—links based on extrasyntactic cues (which can also include phonological cues such as alliteration or rhyme). These mapped elements could then be fed through a process of conceptual combination to create a context-specific semantic representation. This kind of integrated comprehension process has been termed *analogical resonance* (Holyoak & Stamenković, 2018). The couplings created by analogy invite comparisons, which cause word meanings to "resonate" and modify each other. Analogical resonance could underlie poetic effects such as personification (e.g., "The sky weeps" suggests that nature can express a human emotion). This type of subtle meaning adjustment based on the active integration of semantic knowledge is consistent with the longstanding view that metaphor involves the "interanimation of words" (Richards, 1936).

Future research should delve deeper into the comprehension of literary metaphors. The complexities they pose (including emotional as well as strictly cognitive effects) will require integrating behavioral studies with neural and computational investigations (an approach dubbed *neurocognitive poetics* by Jacobs, 2015; see also Holyoak, 2019). Metaphors can be expressed using a wide range of syntactic forms (e.g., Brooke-Rose, 1958; Cardillo et al., 2010), and future studies should examine how the form of linguistic expressions impacts metaphor comprehension. In addition, more research is required to examine the impact on metaphor comprehension of providing a larger pragmatic or linguistic context (e.g., Gibbs & Gerrig, 1989; Ortony, Schallert, Reynolds, & Antos, 1978; Shinjo & Myers, 1987). Finally, more studies should investigate the production as well as comprehension of metaphors (e.g., Chiappe & Chiappe, 2007; Pierce & Chiappe, 2008). The importance of psychological work on metaphor can be expressed by one (adapted from a book title; Handl & Schmid, 2011): metaphor provides a window through which language reveals the human mind.

Acknowledgments

Preparation of this paper was supported by two grants to Dušan Stamenković: Fulbright Visiting Scholar Grant (PS00232724), and a project grant from the Ministry of Education, Science, and Technological Development of the Republic of Serbia (179013); and by NSF Grant BCS-1827374 to Keith Holyoak. A preliminary report of some of this work was presented at the 2018 conference of the Cognitive Science Society. We would like to thank our colleagues and research assistants who helped us in this venture: Airom Bleicher, Vladimir Figar, Lucia Harley, Alexa Hernandez, Kimberly Park, Anshu Patel, Mercan Petek Kuscu, Cristian Ramos, Jocelyn Margarita Reyes, Justin Shin, Angus Chung Hin Tse, Yasemin Yahni, and Jiawen Yu. We are also very grateful to Peter Gordon for guidance in selecting measures of crystalized intelligence, and to Eileen Cardillo for providing us with the full stimuli lists from Cardillo et al. (2010, 2017).

Appendix A

Semantic Similarities Test

How are the two concepts in each pair similar to one another?
For example: "How are a CHAIR and a SOFA similar to one another?"
(ANSWER: Both are types of furniture.)
or: "How are a TURTLE and a TANK similar to one another?"
(ANSWER: Both have a form of armor.)

(1) BIRD – AIRPLANE
(2) SWORD – PISTOL
(3) ORANGE – BALL
(4) SUN – LIGHTBULB
(5) PAPER – LEAF
(6) PEAK – NEEDLE
(7) ROAD – RIVER
(8) LOVE – DRUG
(9) MOUNTAIN – OBSTACLE
(10) CIRCLE – NECKLACE
(11) LONELINESS – DESERT
(12) RIDDLE – LABYRINTH
(13) TIME – RIVER
(14) CORPORATION – TREE
(15) TAVERN – CHURCH
(16) THEORY – BUILDING
(17) DIAMOND – SNOWFLAKE
(18) MEMORY – PRISON
(19) KEY – ANSWER
(20) MARRIAGE – ALLOY

Semantic Similarities Test (scoring key)

(1) bird – airplane (2 pts = fly; 1pt = both have wings, can be seen in the sky)
(2) sword – pistol (2 pts = weapons; 1pt = dangerous, both have a handle, can do damage)
(3) orange – ball (2 pts = round, sphere; 1pt = circle, endless, similar shape)
(4) sun – lightbulb (2 pts = emit/give/provide light, illuminate, source of light; 1pt = bright, shiny)
(5) paper – leaf (2 pts = flat/thin sheet, plant/tree product; 1pt = can be torn, can wrap things, light weight, fragile)
(6) peak – needle (2 pts = pointed tip, pointy/pointed, convex; 1pt = sharp, stick-like, can poke)
(7) road – river (2 pts = transportation paths, pathways, can travel/ride on them; 1pt = both flow, continuous, go for a long stretch, long and narrow, lead somewhere)
(8) love – drug (2 pts = addictive, affect brain/thinking, impair judgment; 1pt = toxic, intoxicating, alter a person, affect actions, give adrenaline, make you do irrational things, provide pleasure)
(9) mountain – obstacle (2 pts = have to/difficult to overcome, impediments/barriers to progress, get in the way; 1pt = take effort, difficult)
(10) circle – necklace (2 pts = round, closed loop, circular, round, ring-like, enclosed; 1pt = similar shape, never ending, go around something, infinite, have an empty space in the middle)
(11) loneliness – desert (2 pts = emptiness, lack of people, desolate, isolation; 1pt = barren, drive someone crazy, boring, vast, sad)
(12) riddle – labyrinth (2 pts = puzzle, hard to find path/solution, can be solved; 1pt = complex, challenging, confusing, tricky, mysterious)
(13) time – river (2 pts = flow in one direction, always passing/moving towards/forward, flow (continuously), keep/constantly moving; 1pt = continuous, cannot be stopped, never ending, running, run)

(14) corporation – tree (2 pts = hierarchical/bottom-up structure, hierarchy, branching; 1pt = both grow, have a strong base/foundation, start of small, can be climbed, have levels)
(15) tavern – church (2 pts = public building, building where people gather/congregate/meet/convene; 1pt = building, structure, shelter, both offer solace/comfort/safety)
(16) theory – building (2 pts = organized structures, include framework, have solid foundations, have to be constructed; 1pt = encompass smaller parts, can be built upon, complex, it takes time to develop them, designed by humans, supported by other things, originate from an idea)
(17) diamond – snowflake (2 pts = crystalline structure, multi-facetted, contain repeated geometric patterns; 1pt = sparkly, shiny, unique, complex, translucent, delicate, products of nature)
(18) memory – prison (2 pts = confining, difficult to escape, limiting, hold captive; 1pt = trap, both store/hold entities, lead to stress)
(19) key – answer (2 pts = achieve goals, open up new possibilities/discoveries; 1pt = both solve a problem, solutions, provide access/means to an end, unlocks)
(20) marriage – alloy (2 pts = blending/melding/bond/fusion/amalgamation/combination of two elements, two elements coming together, stable union; 1pt = combination (and alike w/o two things/elements), malleable, fuse, bonds)

B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2018.12.003.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30–60. https://doi.org/10.1037/0033-2909.131.1.30.

Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven Advanced Progressive Matrices Test. *Journal of Psychoeducational Assessment, 17*, 354–361. https://doi.org/10.1177/073428299901700405.

Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience, 1*, 44–51. https://doi.org/10.1080/17588920903548751.

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage, 55*, 1847–1852. https://doi.org/10.1016/j.neuroimage.2011.01.066.

Black, M. (1962). Metaphor. In M. Black (Ed.). *Models and metaphors* (pp. 38–47). Ithaca, NY: Cornell University Press.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language: A quantitative meta-analysis of neuroimaging studies of metaphor, idiom, and irony processing. *Neuropsychologica, 50*, 2669–2683. https://doi.org/10.1016/j.neuropsychologia.2012.07.021.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*, 193–216. https://doi.org/10.1037/0033-295X.112.1.193.

Brooke-Rose, C. (1958). *A grammar of metaphor.* London: Secker & Warburg.

Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage, 46*, 338–342. https://doi.org/10.1016/j.neuroimage.2009.01.064.

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS ONE, 11*(3), e0150289. https://doi.org/10.1371/journal.pone.0150289.

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavioral Research Methods, 42*, 651–664. https://doi.org/10.3758/BRM.42.3.651.

Cardillo, E. R., Watson, C. E., & Chatterjee, A. (2017). Stimulus needs are a moving target: 240 additional matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavioral Research Methods, 49*, 471–483. https://doi.org/10.3758/s13428-016-0717-1.

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*, 3212–3221. https://doi.org/10.1016/j.neuroimage.2011.11.079.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* New York: Houghton Mifflin.

Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language, 56*, 172–188. https://doi.org/10.1016/j.jml.2006.11.006.

Cho, S., Holyoak, K. J., & Cannon, T. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35*, 1445–1455. https://doi.org/10.3758/BF03193614.

Cho, S., Moody, T. D., Fernandino, L., Mumford, J. A., Poldrack, R. A., Cannon, T. D., ... Holyoak, K. J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex, 20*, 524–533. https://doi.org/10.1093/cercor/bhp121.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behavior. *Trends in Cognitive Sciences, 14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004.

Dunbar, K. N., & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.). *The Oxford handbook of thinking and reasoning* (pp. 701–718). New York: Oxford University Press.

Estes, Z., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory & Cognition, 28*, 28–34. https://doi.org/10.3758/BF03211572.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1–63. https://doi.org/10.1016/0004-3702(89)90077-5.

Gagliano, A., Paul, E., Booten, K., & Hearst, M. A. (2016). Intersecting word vectors to take figurative language to new heights. In Proceedings of the fifth workshop on computational linguistics for literature, NAACL-HLT 2016 (pp. 20–31). San Diego, CA: Association for Computational Linguistics.

Gagné, C. L. (2002). Metaphoric interpretations of comparison-based combinations. *Metaphor and Symbol, 17*, 161–178.

Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Vol. Ed.), *Advances in the Psychology of Learning and Motivation: vol. 22*, (pp. 307–358). New York: Academic Press.

Gernsbacher, M. A., Keysar, B., Robertson, R. R., & Werner, N. K. (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language, 45*, 433–450. https://doi.org/10.1006/jmla.2000.2782.

Gibbs, R. W., Jr., & Gerrig, R. J. (1989). How context makes metaphor comprehension seem 'special'. *Metaphor and Symbolic Activity, 4*, 145–158. https://doi.org/10.1207/s15327868ms0403_3.

Glucksberg, S., & Haught, C. (2006). On the relation between metaphor and simile: When comparison fails. *Mind & Language, 21*, 360–378. https://doi.org/10.1111/j.1468-0017.2006.00282.x.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3–18. https://doi.org/10.1037/0033-295X.97.1.3.

Green, A. E., Kraemer, D. J., Fugelsang, A. J., Gray, J. R., & Dunbar, K. N. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex, 20*, 70–76. https://doi.org/10.1093/cercor/bhp081.

Green, A. E., Kraemer, D. J., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 264–272. https://doi.org/10.1037/a0025764.

Handl, S., & Schmid, H.-. J. (Eds.). (2011). *Windows to the mind: Metaphor, metonymy and conceptual blending.* Berlin: Walter de Gruyter.

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., & Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. *Human Brain Mapping, 37*, 1953–1969. https://doi.org/10.1002/hbm.23149.

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review, 125*, 293–328. https://doi.org/10.1037/rev0000094.

Holyoak, K. J. (1982). An analogical framework for literary interpretation. *Poetics, 11*, 105–126. https://doi.org/10.1016/0304-422X(82)90028-6.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.). *The Oxford handbook of thinking and reasoning* (pp. 234–259). New York: Oxford University Press.

Holyoak, K. J. (2019). *The spider's thread: Metaphor in mind, brain, and poetry.* Cambridge, MA: MIT Press.

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin, 144*, 641–671. https://doi.org/10.1037/bul0000145.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427–466. https://doi.org/10.1037/0033-295X.104.3.427.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220–264. https://doi.org/10.1037/0033-295X.110.2.220.

Jacobs, A. M. (2015). Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers Human*

Neuroscience, 9, 186. https://doi.org/10.3389/fnhum.2015.00186.

Jacobs, A. M., & Kinder, A. (2017). The brain is the prisoner of thought: A machine-learning assisted quantitative narrative analysis of literary metaphors for use in Neurocognitive Poetics. *Metaphor and Symbol, 32*, 139–160. https://doi.org/10.1080/10926488.2017.1338015.

Jacobs, A. M., & Kinder, A. (2018). What makes a metaphor literary? Answers from two computational studies. *Metaphor and Symbol, 33*, 85–100. https://doi.org/10.1080/10926488.2018.1434943.

Jones, L. L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language, 53*, 110–124. https://doi.org/10.1016/j.jml.2005.01.016.

Katz, A., Paivio, A., Marschark, M., & Clark, J. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity, 3*, 191–214. https://doi.org/10.1207/s15327868ms0304_1.

Kazmerski, V. A., Blasko, D. G., & Dessalegn, B. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory & Cognition, 31*, 673–689. https://doi.org/10.3758/BF03196107.

Keane, M. T., & Costello, F. J. (2001). Setting limits on analogy: Why conceptual combination is not structural alignment. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.). *The analogical mind: Perspectives from cognitive science* (pp. 287–312). Cambridge, MA: MIT Press.

Kertész, A., Rákosi, C., & Csatár, P. (2012). Data, problems, heuristics and results in cognitive metaphor research. *Language Sciences, 34*, 715–727. https://doi.org/10.1016/j.langsci.2012.04.011.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*, 257–266. https://doi.org/10.3758/BF03212981.

Kintsch, W. (2001). Predication. *Cognitive Science, 25*, 173–202. https://doi.org/10.1207/s15516709cog2502_1.

Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol, 17*, 249–262. https://doi.org/10.1207/S15327868MS1704_1.

Kubricht, J. R., Lu, H., & Holyoak, K. J. (2017). Individual differences in spontaneous analogical transfer. *Memory & Cognition, 45*, 576–588. https://doi.org/10.3758/s13421-016-0687-7.

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

Levin, S. R. (1962). *Linguistic structures in poetry.* The Hague, Netherlands: Mouton.

McGlone, M. S., & Manfredi, D. A. (2001). Topic-vehicle interaction in metaphor comprehension. *Memory & Cognition, 29*, 1209–1219. https://doi.org/10.3758/BF03206390.

Morrison, R. G., Krawczyk, D., Holyoak, K. J., Hummel, J. E., Chow, T., Miller, B., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal dementia. *Journal of Cognitive Neuroscience, 16*, 260–271. https://doi.org/10.1162/089892904322984553.

Newsome, M. R., & Glucksberg, S. (2002). Older adults filter irrelevant information during metaphor comprehension. *Experimental Aging Research, 28*, 253–267. https://doi.org/10.1080/03610730290080317.

Nippold, M. A., & Sullivan, M. P. (1987). Verbal and perceptual analogical reasoning and proportional metaphor comprehension. *Journal of Speech and Hearing Research, 30*, 367–376. https://doi.org/10.1044/jshr.3003.367.

Olkoniemi, H., Ranta, H., & Kaakinen, J. H. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 433–450. https://doi.org/10.1037/xlm0000176.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review, 86*, 161–180. https://doi.org/10.1037/0033-295X.86.3.161.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. L. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17*, 465–477. https://doi.org/10.1016/S0022-5371(78)90283-9.

Patterson, K. J. (2016). The analysis of metaphor: To what extent can the theory of lexical priming help our understanding of metaphor usage and comprehension? *Journal of Psycholinguistic Research, 45*, 237–258. https://doi.org/10.1007/s10936-014-9343-1.

Pierce, R. S., & Chiappe, D. L. (2008). The roles of aptness, conventionality, and working memory in the production of metaphors and similes. *Metaphor and Symbol, 24*, 1–19. https://doi.org/10.1080/10926480802568422.

Prat, C. S., Mason, R. A., & Just, M. A. (2012). An fMRI investigation of analogical mapping in metaphor comprehension: The influence of context and individual cognitive capacities on processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 282–294. https://doi.org/10.1037/a0026037.

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *Neuroimage, 62*, 600–610. https://doi.org/10.1016/j.neuroimage.2012.06.022.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence, 1938, individual form.* London: Lewis.

Richards, I. A. (1936). *The philosophy of rhetoric.* New York: Oxford University Press.

Shinjo, M., & Myers, J. L. (1987). The role of context in metaphor comprehension. *Journal of Memory and Language, 26*, 226–241. https://doi.org/10.1016/0749-596X(87)90125-2.

Snow, R. E., Kyllonen, C. P., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Vol. Ed.), *Advances in the psychology of human intelligence: vol. 2*, (pp. 47–103). Hillsdale, NJ: Erlbaum.

Spearman, C. (1927). *The abilities of man.* New York: MacMillan.

Tourangeau, R., & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language, 30*, 452–472. https://doi.org/10.1016/0749-596X(91)90016-D.

Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology, 13*, 27–55. https://doi.org/10.1016/0010-0285(81)90003-7.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203–244. https://doi.org/10.1016/0010-0277(82)90016-6.

Trick, L., & Katz, A. (1986). The domain interaction approach to metaphor processing: Relating individual differences and metaphor characteristics. *Metaphor and Symbolic Activity, 1*, 185–213. https://doi.org/10.1207/s15327868ms0103_3.

Vartanian, O. (2012). Dissociable neural systems for analogy and metaphor: Implications for the neuroscience of creativity. *British Journal of Psychology, 103*, 302–316. https://doi.org/10.1111/j.2044-8295.2011.02073.x.

Vendetti, M., Wu, A., & Holyoak, K. J. (2014). Far out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science, 25*, 928–933. https://doi.org/10.1177/0956797613518079.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition and analogical reasoning in older adults. *Psychology and Aging, 19*, 581–591. https://doi.org/10.1037/0882-7974.19.4.581.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., ... Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125. https://doi.org/10.1111/1467-9280.00118.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28*, 1205–1212. https://doi.org/10.3758/BF03211821.

Wechsler, D. (1997). *WAIS-III: Administration and scoring manual – Wechsler adult intelligence scale.* London: The Psychological Corporation/Pearson Education.

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4*, 167–183.

Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language, 38*, 177–202.