

A 25-year replication of Katz et al.'s (1988) metaphor norms

Spencer J. Campbell · Gary E. Raney

Published online: 14 March 2015
© Psychonomic Society, Inc. 2015

Abstract Research in metaphor processing has made extensive use of the normed metaphor database created by Katz, Paivio, Marschark, & Clark (*Metaphor and Symbolic Activity*, 3, 191–214, 1988). Because of the plasticity of figurative language, we conducted a renorming of selected metaphors from the database on a new student population. Correlations between Katz et al.'s and the present data showed that the pattern of responses has remained highly consistent across time and populations. The consistency of the normative ratings allows us to be confident in future research that will use the Katz et al. collection.

Keywords Metaphor · Figurative language · Norming · Individual differences · Katz

The use of controlled stimuli is an essential component of the scientific process, so it is important to ensure that stimuli have been appropriately normed for the population and variables being tested. Oftentimes researchers will use a shared database of normed stimuli to ensure consistency across projects and laboratories. One such collection of normed stimuli is the set of literary and nonliterary metaphors generated by Katz, Paivio, Marschark, and Clark (1988). Katz et al. collected ratings on ten dimensions that could be used to describe metaphors: their comprehensibility, ease of interpretation, metaphoricity, metaphor goodness, imagery of the metaphor, imagery of the subject, imagery of the predicate, familiarity, semantic relatedness, and number of alternative interpretations.

Because of the span of dimensions measured, this collection has been used in many studies since its publication (e.g., Bowdle & Gentner, 2005; Campbell & Raney, 2013; De Grauwe, Swain, Holcomb, Ditman, & Kuperberg, 2010; Diaz, Barrett, & Hogstrom, 2011; Diaz & Hogstrom, 2011; Gentner & Wolff, 1997; Kacirik & Chiarello, 2007; Kuiken, Chudleigh, & Racher, 2010; Schmidt & Seger, 2009; Thibodeau, & Durgin, 2011; Xu, 2010). The purpose of the present research was to replicate a portion of Katz et al.'s metaphor norms to determine whether their normative data are still valid.

There are several reasons to replicate Katz et al.'s (1988) norms. One important reason is that language is not stagnant, and interpretations of various figurative tropes could change over time. Therefore, it is reasonable to ask whether the metaphors used by Katz et al. are perceived and understood in the same way today as they were 25 years ago. For instance, the familiarity of the metaphors may have changed, or the conventionality of the base may have shifted over time (Bowdle & Gentner, 2005)—where the *base* refers to the final word of a metaphor (the base is also commonly referred to as the vehicle). For example, many conventional words, such as *gold mine* or *blockbuster*, are entrenched in our lexicon as figurative phrases, largely ignoring their literal origin.

Another reason to replicate Katz et al.'s (1988) norms is that different groups of people might respond to metaphors differently. As is typical of much research in psychology, Katz et al. used undergraduate students as participants, but note that the research was performed at a university in Canada. Although we have no reason to expect that Canadian students would process metaphors much differently than say, students in the United States, there is evidence that interpretations of metaphors may differ between cultures or geographic regions (Boers, 2003; Littlemore & Low, 2006). For instance, when the word *ski* is presented in an ambiguous context (e.g., *I want to go skiing*) to people who live in the state of Florida, their initial thought might be of water skiing,

S. J. Campbell (✉) · G. E. Raney
Department of Psychology (MC 285), University of Illinois at
Chicago, Chicago, IL 60607, USA
e-mail: scampb21@uic.edu

G. E. Raney
e-mail: geraney@uic.edu

whereas people who live in Wyoming might initially think of snow skiing. Consequently, it is important to determine whether the normative ratings would replicate using a different population. Beyond knowing that Katz et al.'s participants were college students in introductory psychology, little is known about the participants. As Katz et al. pointed out, differences between individuals may impact the characteristics of the metaphor that are readily available to the reader. What sorts of factors may elicit these differences in ratings? One potential factor is the language backgrounds of the sample population used. For instance, whether normative ratings are collected using predominantly monolingual or bilingual participants might be important, because monolinguals and bilinguals have different linguistic and perhaps cultural experiences (Bortfeld, 2002; Colston, 2005).

Because the database created by Katz et al. (1988) has been and continues to be used extensively in research on metaphor processing, we believe that the norming data should be compared to data from a new generation of research participants, to examine the modern validity of the original norms. What follows is a renorming of a sample of the nonliterary metaphors included in Katz et al.'s study.

Methodology

Participants

Ninety undergraduate students from the University of Illinois at Chicago (UIC) participated in exchange for credit toward their introductory psychology course.¹ UIC's student population represents one of the most diverse campuses in the United States, in that it is a minority-majority campus. This means that no one racial group comprises at least half of the total student population. UIC is also linguistically diverse. Approximately 52 % of UIC students self-report being multilingual, and another 29 % report being "somewhat" multilingual (i.e., not proficient in their second language). Furthermore, approximately 40 % of UIC students indicate that English is not their native language or that they have two native languages (i.e., they learned two languages from birth or from shortly after birth), but the vast majority of students had attended English-speaking schools prior to college, making them highly proficient English speakers.

Because of the diverse language background of the students at UIC, participants were required to have attended English-speaking schools for at least 10 years. This restriction was used to ensure that they had substantial knowledge of the

English language. Of the 90 students tested, 60 self-reported being proficient bilinguals (66 %), and 60 (66 %) were native English speakers. This information was gathered through self-report language history questionnaires. Descriptive information about the participants is provided in Table 1.

Materials and apparatus

Fifty nonliterary metaphors were selected from the Katz et al. (1988) collection. Because this collection is oftentimes used for research exploring familiarity effects on metaphor processing, the metaphors were selected from a wide range of Katz et al.'s familiarity ratings. Specifically, 20 of the most familiar metaphors, 20 of the most unfamiliar metaphors, and 10 metaphors near the median familiarity score were selected. Additionally, given that syntactic structure can have a substantial impact on comprehension (Gentner & Wolff, 1997; Glucksberg, 2008), all of the selected metaphors followed the "X is a Y" format. For example, the metaphor *Love is a flower* follows this format, whereas *Thunder clouds are draperies pulled across the sun* does not. A full list of the metaphors used can be found in Appendix B.

The 50 selected metaphors were printed in packets with the metaphors in a predetermined, randomized order. Each metaphor was presented alone (i.e., with no context) and was followed by the ten norming questions Katz et al. (1988) had used to evaluate (1) comprehensibility, (2) ease of interpretation, (3) metaphoricity, (4) metaphor goodness, (5) metaphor imagery, (6) subject imagery, (7) predicate imagery, (8) felt familiarity, (9) semantic relatedness, and (10) number of

Table 1 Average scores (and standard deviations in parentheses) on the vocabulary test (maximum = 30) and self-reported proficiency ratings (maximum = 10) for speaking, understanding, and reading English and for participants' most proficient language other than English

Measure	Overall	Group 1	Group 2	Group 3
Vocabulary	14.3 (4.0)	14.4 (4.2)	14.3 (4.2)	14.0 (3.9)
English				
Speak	9.5 (1.0)	9.6 (0.8)	9.6 (0.9)	9.3 (1.4)
Understand	9.4 (1.1)	9.6 (0.7)	9.4 (1.0)	9.3 (1.5)
Reading	9.4 (1.1)	9.6 (0.8)	9.4 (1.1)	9.1 (1.4)
Most Proficient Language Other Than English				
Speak	6.4 (2.7)	3.2 (2.0)	7.1 (1.1)	8.1 (2.0)
Understand	6.8 (2.7)	3.5 (2.1)	7.5 (1.6)	8.5 (1.8)
Read	5.2 (3.2)	3.2 (2.6)	5.7 (2.8)	6.3 (3.2)

"Overall" represents the average across all participants. Participants were also divided into one of three groups: Group 1 included native English speakers who were not proficient in a second language (L2), Group 2 included native English speakers who were proficient in an L2, and Group 3 included nonnative English speakers who were proficient in English.

¹ Means were calculated after running approximately 30 participants and then recalculated after tripling the sample. The means were highly similar for the first and second halves of the participants; therefore, the data were deemed stable, and data collection was stopped.

interpretations. See [Appendix A](#) for a brief description of each of these dimensions from the instruction page of the norming packet, and Katz et al.'s report for a full explanation. The descriptions used for our norming packet were taken from Katz et al., and each domain was rated on a seven-point scale, with each scale being explained to the participants before they began the study.

As part of the metaphor norming packet, participants were given a language history questionnaire and an English vocabulary test (developed by Raney). The language history questionnaire allowed us to collect self-report information on the participants' native languages, the number of languages they speak, and their relative strengths using each language. The vocabulary test had been used in a number of prior studies (Minkoff & Raney, 2000; Therriault & Raney, 2007) and is moderately correlated with English reading comprehension ability ($r_s = .40$ to $.52$ in prior studies). The test consisted of 30 words presented in isolation, and the participants were asked to select the best meaning from among five alternatives. The vocabulary test was designed to be relatively difficult, with the average score being approximately 14–15 correct for a population of predominantly freshman college students.

Procedure

Participants were provided with a metaphor norming packet containing instructions, the metaphors, the language history questionnaire, and the vocabulary quiz, in that order. Each of the norming dimensions was described to the participants before they began rating the metaphors. This was particularly important for clarifying the subject and predicate imagery dimensions (the subject is often called the *target*, and the predicate is often called the *vehicle* or *base*, in metaphor research). Participants were instructed to complete the norming packet before completing the language history questionnaire and vocabulary test. After completing all three forms, the participants were debriefed and dismissed.

Results

We present the results in two sections. The first section reports overall comparisons between the Katz et al. (1988) and UIC ratings. Interscale correlations between the ten dimensions are then provided for the UIC data (see Katz et al., 1988, for their interscale correlations). The second section reports comparisons of the Katz et al. ratings and the UIC ratings when the UIC participants were divided into subgroups on the basis of their vocabulary scores and language background.

Overall comparisons

Ratings and correlations Average ratings for each of the ten norming dimensions were collected for each metaphor and then correlated with the ratings collected by Katz et al. (1988), using Pearson correlations. Table 2 presents the average ratings for the UIC and Katz et al. participants. The ratings of the selected metaphors have remained highly consistent over time, as is indicated by the presence of significant positive correlations between the UIC and Katz et al. ratings for all ten dimensions. The correlations range from $.56$ to $.78$, with six of the ten correlations exceeding $.70$. Of particular importance is the high correlation for the familiarity dimension ($r = .78$), which indicates that the metaphors that were rated as familiar 25 years ago are still considered familiar today, and the metaphors that were rated as unfamiliar have remained relatively unfamiliar.

The absolute ratings have increased slightly over time for some of the dimensions. The important finding, however, is that the relative ratings have remained stable, as is indicated by the correlations between the Katz et al. (1988) and UIC ratings. Researchers typically use normative ratings to access metaphors that are relatively high or low in familiarity, such as the top or bottom quartiles. The absolute values of the ratings are secondary to the relative ratings.

Difference scores Although the relative ratings across dimensions have remained consistent, Table 2 shows that the magnitudes of the average ratings have shifted slightly over time. Specifically, the average ratings for eight of the ten dimensions are slightly larger for the UIC population than for the

Table 2 Average ratings (maximum = 7, with standard deviations in parentheses) for each metaphor dimension, correlations between the Katz et al. (1988) and UIC ratings, and difference scores between the Katz et al. and UIC ratings

Dimension	UIC	Katz et al.	Correlation	Difference
Comprehensibility	4.7 (1.0)	4.5 (1.0)	.71***	0.23
Ease of interpretation	4.6 (1.1)	4.6 (1.1)	.74***	-0.05
Metaphoricity	4.3 (1.1)	2.9 (0.7)	.56***	1.44**
Metaphor goodness	4.2 (1.1)	3.6 (0.9)	.65***	0.62**
Metaphor imagery	4.5 (1.1)	3.5 (1.1)	.72***	1.0***
Subject imagery	4.6(0.9)	3.7 (1.0)	.75***	0.94***
Predicate imagery	4.5 (0.9)	3.7 (1.0)	.59***	0.80***
Felt familiarity	4.0 (1.0)	3.3 (1.3)	.78***	0.74**
Semantic relatedness	3.7 (1.0)	3.4 (1.1)	.75***	0.33
Interpretations	1.1 (0.4)	1.6 (0.4)	.68***	-0.49***

Difference scores were calculated by the subtracting the Katz et al. ratings from the UIC ratings, with the absolute difference being reported. Difference scores were compared to zero. Significant correlations and difference scores are indicated by asterisks ($^* p < .05$, $^{**} p < .01$, $^{***} p < .001$).

Katz et al. (1988) population. Ratings for two of the dimensions (metaphoricity and metaphor imagery) have increased by over one point (one point equals approximately a 14 % change). To evaluate the change in ratings, independent-samples *t* tests between the average UIC ratings and the Katz et al. ratings were run for each dimension (see Table 3). These tests were based on the average rating for each dimension for each participant (i.e., one score per participant per dimension). With the exception of comprehensibility, ease of interpretation, and semantic relatedness (*ts* < 2.0, n.s.), the ratings from UIC were significantly higher than the Katz et al. norms (all *ps* < .05).

Interscale correlations These correlations were computed for each of the ten scales to determine how related they are to each other, as had been done by Katz et al. (1988). Table 4 shows that all of the dimensions are strongly correlated with one another. It is worth noting that the interscale correlations found with the UIC population are larger than those collected by Katz et al. The average interscale correlation reported by Katz et al. was .76, whereas the UIC average is .94.

There are two potential reasons for the higher interscale correlations for the UIC data. The first reason is that the UIC data are based only on nonliterary metaphors in the “X is a Y” format, in which X and Y are single words. Thus, the type of metaphors evaluated for the UIC norms was more restricted than the types of metaphors included in Katz et al.’s (1988) norms. To evaluate this possibility, we recalculated the interscale correlations for Katz et al.’s norms on the basis of the 50 metaphors included in the present study. The average interscale correlation increased from .76 (Katz et al.’s, 1988, original average) to .81. Restricting the metaphors led to a small increase in the average interscale correlation, but certainly not to a point equal to our average

interscale correlation. This makes sense, given that the 50 ratings were taken from a database in which participants rated the full set of 260 nonliterary metaphors; thus, calculating interscale correlations on the subset of 50 items does not minimize the range of metaphors actually rated by Katz et al.’s participants.

A second potential reason for the difference in the interscale correlations is that Katz et al. (1988) had separate groups of participants rate all of the metaphors on a single dimension, rather than having a single group of participants rate the metaphors on all ten dimensions (as we did). To evaluate whether this methodological change influenced the size of the interscale correlations, we conducted a follow-up experiment in which participants rated the metaphors on a single dimension. We included three dimensions (comprehensibility, familiarity, and number of alternative interpretations). The follow-up study is reported fully in Appendix C. The key finding from the follow-up study is that changing the procedure did not substantially influence the results. The average correlations between the UIC data and Kata et al.’s data were .67, .82, and .62 for comprehensibility, familiarity, and number of alternative interpretations, respectively. The average interscale correlation between these three dimensions was approximately .9. Thus, for these dimensions, asking participants to rate all of the dimensions or to rate a single dimension did not change the pattern of results. This does not explain why the interscale correlations are high for the UIC data, but this eliminates the methodological explanation.

Individual differences

The UIC metaphor ratings were reevaluated to determine whether individual differences in English vocabulary knowledge and language history influenced the ratings. Specifically, participants were divided into low- and high-vocabulary groups and into three language groups, based on whether they were native or nonnative English speakers and whether the native English speakers were proficient or nonproficient bilinguals.

Vocabulary The participants were divided into low and high vocabulary knowledge groups based on a median split of their vocabulary scores. Low-vocabulary participants scored 13 points or lower out of a possible 30 on the vocabulary test, with the average score being 11.2. The average score for the high-vocabulary group was 17.8. A *t* test showed that the difference in vocabulary scores between the groups was significant, $t(88) = 10.6$, $p < .01$.

The average ratings for both vocabulary groups for each of the ten norming dimensions, as well as the correlations with the Katz et al. (1988) ratings, can be found in Table 5. For both the low- and high-vocabulary groups, we found significant correlations between the UIC ratings and Katz et al.’s ratings.

Table 3 Independent-samples *t* tests comparing average UIC ratings and the average ratings collected by Katz et al. (1988) for each norming dimension

Dimension	<i>t</i> Test	<i>p</i> Value
Comprehensibility	$t(98) = 1.2$	$p = .23$
Ease of interpretation	$t(98) = -0.24$	$p = .81$
Metaphoricity	$t(98) = 8.0$	$p < .01$
Metaphor goodness	$t(98) = 3.2$	$p < .01$
Metaphor imagery	$t(98) = 4.8$	$p < .01$
Subject imagery	$t(98) = 5.1$	$p < .01$
Predicate imagery	$t(98) = 4.2$	$p < .01$
Felt familiarity	$t(98) = 3.3$	$p < .01$
Semantic relatedness	$t(98) = 1.6$	$p = .12$
Interpretations	$t(98) = -6.5$	$p < .01$

Table 4 Interscale correlations between metaphor dimensions

	CMP	ESI	MET	MGD	IMG	IMS	IMP	FAM	SRL	ALT
CMP	1	.99*	.99*	.97*	.98*	.94*	.96*	.97*	.97*	.87*
ESI		1	.99*	.98*	.98*	.95*	.96*	.98*	.98*	.85*
MET			1	.98*	.97*	.94*	.95*	.98*	.98*	.85*
MGD				1	.96*	.94*	.95*	.98*	.97*	.86*
IMG					1	.98*	.98*	.97*	.97*	.85*
IMS						1	.97*	.95*	.95*	.83*
IMP							1	.96*	.95*	.84*
FAM								1	.99*	.85*
SRL									1	.83*
ALT										1

All correlations are significant (*) at $p < .001$. CMP = comprehensibility, ESI = ease of interpretation, MET = metaphoricity, MGD = metaphor goodness, IMG = metaphor imagery, IMS = subject imagery, IMP = predicate imagery, FAM = familiarity, SRL = semantic relatedness, ALT = number of alternative interpretations.

Examination of the ratings in Table 5 gives the impression that the average ratings for most dimensions are slightly higher for the low-vocabulary than for the high-vocabulary group. To evaluate this possibility, we compared the average ratings, collapsed across the ten dimensions, for the low- and high-vocabulary participants (i.e., one data point per participant to represent the average rating across all ten dimensions). Average ratings were not significantly different between the vocabulary groups [$t(998) = 0.14$, n.s.].

Most importantly, the correlations between each vocabulary group and the Katz et al. (1988) ratings were large and statistically significant for every dimension. Participants in the low-vocabulary group appear to have smaller correlations

than the high-vocabulary group for nearly every dimension, but the average correlation across the ten dimensions was not statistically lower for the low-vocabulary group, $t(18) = -1.6$, n.s. In essence, vocabulary knowledge did not significantly affect the magnitude of the UIC ratings or the correlations with the Katz et al. ratings.

Language history Participants were split into three groups based on their responses on the language history questionnaire. Group 1 consisted of native English speakers who were nonproficient bilinguals ($n = 30$). Most of these individuals had some experience with a second language, usually from learning it in a classroom, but did not consider themselves proficient in speaking, understanding, or reading the language. Only four individuals in Group 1 considered themselves purely monolingual, with no experience learning a foreign language. Group 2 consisted of native English speakers who were proficient bilinguals ($n = 30$). These individuals reported themselves as having first learned English, but they were also capable of proficiently speaking a second language (usually learned at an early age). The second language was often learned in a school setting as well as at home. Group 3 consisted of nonnative English speakers who were proficient bilinguals ($n = 30$). The individuals were proficient in English and in another language that had been learned prior to acquiring English, which was typically their native language. The second language was often learned in a school setting as well as at home.

Table 6 presents the average ratings for the three UIC language groups, as well as the correlations between the ratings from each language group and the Katz et al. (1988) participants. Across all ten dimensions, the ratings for each language group were significantly correlated with the Katz et al. ratings. A one-way between-subjects analysis of variance (ANOVA)

Table 5 Average ratings (maximum = 7, with standard deviations in parentheses) for low- and high-vocabulary UIC students, and correlations between the Katz et al. (1988) and UIC ratings based on vocabulary score

Dimension	Ratings		Correlations	
	Low Vocab.	High Vocab.	Low Vocab.	High Vocab.
Comprehensibility	4.6 (0.8)	4.9 (1.0)	.68*	.72*
Ease of interpretation	4.7 (1.0)	4.4 (1.2)	.73*	.73*
Metaphoricity	4.4 (1.0)	4.2 (1.1)	.52*	.58*
Metaphor goodness	4.2 (1.0)	4.1 (1.1)	.59*	.69*
Metaphor imagery	4.6 (1.0)	4.3 (1.1)	.70*	.72*
Subject imagery	4.8 (0.8)	4.4 (0.9)	.72*	.75*
Predicate imagery	4.7 (0.9)	4.3 (0.9)	.56*	.61*
Felt familiarity	4.2 (1.0)	3.8 (1.1)	.74*	.79*
Semantic relatedness	3.9 (1.0)	3.5 (1.0)	.70*	.78*
Interpretations	1.3 (0.4)	.94 (0.3)	.62*	.70*

All correlations are significant (*) at $p < .001$.

Table 6 Average ratings (maximum = 7, with standard deviations in parentheses) for UIC students as a function of their language background, and correlations between the Katz et al. (1988) and UIC ratings based on the language background of the UIC students

Dimension	Ratings			Correlations		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
CMP	4.3 (0.9)	5.0 (0.9)	4.8 (1.0)	.72*	.71*	.65*
ESI	4.5 (1.2)	4.5 (1.1)	4.6 (1.1)	.76*	.73*	.69*
MET	4.2 (1.2)	4.3 (1.1)	4.4 (1.0)	.60*	.53*	.49*
MGD	4.2 (1.1)	4.1 (1.1)	4.2 (1.0)	.70*	.61*	.56*
IMG	4.4 (1.1)	4.4 (1.1)	4.7 (1.1)	.73*	.73*	.63*
IMS	4.6 (1.0)	4.5 (1.0)	4.9 (0.8)	.77*	.72*	.66*
IMP	4.5 (1.0)	4.3 (1.0)	4.8 (0.8)	.60*	.60*	.51*
FAM	3.9 (1.0)	3.9 (1.1)	4.3 (1.0)	.82*	.75*	.69*
SRL	3.8 (1.1)	3.6 (1.1)	3.9 (1.0)	.79*	.70*	.68*
ALT	1.5 (0.4)	1.2 (0.4)	1.2 (0.4)	.66*	.54*	.68*

Group 1 included native English speakers who were not proficient in a second language (L2), Group 2 included native English speakers who were proficient in an L2, and Group 3 included nonnative English speakers who were proficient in English. CMP = comprehensibility, ESI = ease of interpretation, MET = metaphoricity, MGD = metaphor goodness, IMG = metaphor imagery, IMS = subject imagery, IMP = predicate imagery, FAM = familiarity, SRL = semantic relatedness, ALT = number of alternative interpretations. All correlations are significant (*) at $p < .001$.

was performed on the average ratings for the three language groups (with the average rating across all ten dimensions for each participant being entered as a data point) to determine whether the average ratings differed across language groups. The average ratings across the ten dimensions were not significantly different between the three language groups, $F(2, 1497) = 1.7$, n.s.

The average correlations between each language group and the Katz et al. (1988) data were also compared using a one-way, between-subjects ANOVA (based on one overall average correlation per participant). The average correlation was significantly different between groups, $F(2, 27) = 3.4$, $p < .05$. Post hoc comparisons showed that Group 1 ($M = .72$) had a statistically larger average correlation with the Katz et al. data than did Group 3 ($M = .63$). The sizes of the correlations with the Katz et al. data did not differ between Groups 2 ($M = .66$) and 3 ($M = .63$).

General discussion

Our findings support the conclusion that the Katz et al. (1988) norms for nonliterary metaphors of the “X is a Y” format remain highly valid. We found significant

correlations between the UIC and Katz ratings for all ten dimensions. Although the magnitudes of the ratings for our participants were slightly larger for some of the dimensions, relative familiarity, meaningfulness, and so forth have remained highly consistent for the metaphors normed here. That is, metaphors that were rated as more familiar 25 years ago were still rated as more familiar now, and metaphors that were rated as less familiar 25 years ago were still rated as less familiar now. The same conclusion holds for the other dimensions. In general, the metaphors in the UIC sample are as comprehensible and as easy or difficult to interpret now as when they were originally normed. Likewise, individuals are generally able to create the same number of interpretations as before, the degree of imagery invoked by the metaphors and their components has remained consistent, and so forth. As such, we can be confident in past and future research that is based on this metaphor collection. The consistency of the familiarity ratings is especially important, because metaphor familiarity is a central component of several models of metaphor processing and has been the focus of much research (Blasko & Briehl, 1997; Blasko & Connine, 1993; Bowdle & Gentner, 2005; Campbell & Raney, 2013; Gentner & Wolff, 1997).

One might wonder why the ratings were slightly higher for some of the dimensions in the UIC data than in the Katz et al. (1988) data. One possible explanation for the higher metaphoricity and familiarity ratings is that metaphors may be encountered more frequently today than 25 years ago. For instance, with increased interactions through brief e-mails and text messaging, people may more frequently use figurative language to quickly express themselves. It is also possible that students might rate any form of language as being more familiar today, and the higher ratings may have nothing to do with knowledge of metaphors per se. These explanations are purely speculative and warrant further study. The important point is that the relative ratings have remained stable over time, as reflected by the strong correlations between the UIC and Katz et al. data. This allows us to be confident in past and current research based on the metaphors in the Katz et al. norms.

The present data also have implications for research based on diverse populations. Large and statistically reliable correlations were found between the UIC ratings and the Katz et al. (1988) ratings, based both on all participants and on the participants divided into groups based on vocabulary knowledge and language background. Vocabulary knowledge, which is moderately correlated with reading comprehension ability, did not affect the magnitude of the ratings or the size and

pattern of the correlations between the UIC and Katz et al. ratings. This implies that the norms are valid for college students of all vocabulary levels, as long as they are proficient speakers of English (defined in the present study as having at least 10 years of education in a setting in which English has been spoken).

Our results also demonstrate that the Katz et al. (1988) database is valid for both native and nonnative English speakers, as long as they are proficient speakers of English. We found large and statistically reliable correlations between the UIC and Katz et al. ratings for native English speakers who were not proficient bilinguals (Group 1), for native English speakers who were proficient bilinguals (Group 2), and for nonnative English speakers who were proficient bilinguals (Group 3). Interestingly, Group 1 consistently had the largest correlations with the Katz et al. data. The larger correlations might be due to the fact that these individuals speak English almost exclusively, and therefore they might have more experience with the selected metaphors. The two bilingual language groups (2 and 3) produced similar ratings and correlations across all ten dimensions. This might be due to the fact that these individuals regularly use multiple languages, and their levels of exposure to the selected metaphors are therefore similar. Future research could examine these speculative explanations regarding exposure to metaphors.

One general implication of our findings is that how linguistic background influences performance is complex. Linguistic background had little effect on the relative ratings of metaphors, such as their relative familiarity. In contrast, it had a reliable effect on the magnitude of ratings, with bilinguals generally rating metaphors as less familiar, for example. This pattern held for both native and nonnative English speakers. How language background influences the comprehension of metaphors remains an important topic for future research.

Some potential limitations need to be considered regarding the present study. First, we used a methodology modified from the one used by Katz et al. (1988). As we mentioned earlier, Katz et al. had separate groups of participants rate the metaphors on a single dimension, whereas we had a single group of participants rate the metaphors on all ten dimensions. Our follow-up experiment (see Appendix C) indicated that the same patterns of results were found using each methodology; therefore, we are confident that the relative ratings remain consistent no matter which methodology is employed. A second potential limitation is that we used a subset of the 264 metaphors normed by Katz et al.—specifically; nonliterary metaphors having the “X is a Y” format, in which X and Y are single words. This could reduce the variability in the ratings relative to rating metaphors

of mixed formats. Future studies could evaluate this possibility.

In summary, our findings support the conclusion that Katz et al.’s (1988) normative ratings of literary metaphors remain valid as long as the research participants are proficient English speakers. Whether a participant has a low or high vocabulary or is a native or nonnative English speaker does not impact the pattern of ratings for this collection of metaphors.

Appendix A: Instruction page for norming packet and sample stimuli with scales

Instructions: You will read sentences and rate them on a number of different categories, using a scale of 1–7, with 1 being the lowest rating of a category and 7 being the highest. The categories you will be rating the sentences on are:

Comprehensibility: How easily are you able to comprehend the sentence?

Ease of Interpretation: How easily can you interpret the meaning of the sentence?

Metaphoricity: To what degree is the sentence figuratively true?

Metaphor Goodness: How good, apt, and pleasing is the metaphor?

Metaphor Imagery: How easily did the sentence create mental images such as mental pictures and/or sounds?

Subject Imagery: Rank the imagery of the subject of the sentence.

Predicate Imagery: Rank the imagery of the predicate of the sentence.

Felt Familiarity: How familiar or how frequently are the ideas expressed in the sentence?

Semantic Relatedness: How related or similarly related are the two items being compared in the sentence?

Number of Alternative Interpretations: How many different interpretations can you think of for the sentence?

Example metaphor with rating scales from norming packet

1. Loneliness is a desert.

Comprehensibility:	1	2	3	4	5	6	7
Ease of Interpretation:	1	2	3	4	5	6	7
Metaphoricity:	1	2	3	4	5	6	7
Metaphor Goodness:	1	2	3	4	5	6	7
Metaphor Imagery:	1	2	3	4	5	6	7
Subject Imagery:	1	2	3	4	5	6	7
Predicate Imagery:	1	2	3	4	5	6	7
Felt Familiarity:	1	2	3	4	5	6	7
Semantic Relatedness:	1	2	3	4	5	6	7
Number of Alternative Interpretations:	_____						

Appendix B

Table 7 Average scores on each norming dimension for each metaphor

CMP	ESI	MET	MGD	IMG	IMS	IMP	FAM	SRL	ALT	
1. Wound is a fjord	2.4/4.6	2.0/4.7	1.9/2.9	1.8/3.5	1.9/3.8	2.6/3.6	2.1/3.8	1.8/2.2	1.7/3.2	0.5/1.2
2. Sailboat is a cat	3.0/2.6	2.6/3.0	2.1/2.1	2.0/2.0	2.3/1.9	3.1/2.2	3.1/2.3	2.1/1.6	2.0/1.9	0.5/1.4
3. Braggart is a snowman	2.9/3.1	2.4/2.8	2.5/2.6	2.3/2.3	2.6/1.9	2.7/2.6	3.0/2.6	2.2/1.8	2.0/2.1	0.4/0.9
4. Lie is a shoal	3.5/3.8	3.0/3.6	3.0/2.9	2.9/3.4	3.0/2.4	3.2/3.4	3.2/2.6	2.7/2.4	2.4/3.0	0.7/0.8
5. Creativity is a toaster	3.6/2.5	3.1/3.0	2.9/2.4	2.9/2.4	3.1/1.9	3.4/2.0	3.5/2.9	2.8/1.5	2.6/2.1	0.8/1.2
6. Island is a cork	3.9/4.5	3.1/4.5	3.1/2.2	2.9/3.3	3.4/3.6	3.6/3.9	3.6/3.9	2.8/2.5	2.4/3.0	1.1/1.4
7. Wind is a cat	3.8/2.9	3.2/2.8	3.0/2.0	3.0/2.1	3.4/1.5	3.7/1.9	3.6/2.4	3.1/1.7	2.6/1.6	1.2/1.3
8. Silence is an apron	3.8/3.3	3.5/3.7	3.3/2.5	3.3/2.3	3.5/2.0	4.0/2.7	4.0/3.1	3.1/2.3	2.8/2.5	0.9/1.2
9. Humor is a salve	3.9/3.5	3.5/4.0	3.5/2.9	3.5/3.6	3.4/2.3	3.7/3.1	3.5/3.1	3.1/2.9	3.0/3.3	0.9/1.2
10. Fisherman is a spider	4.3/3.6	3.9/3.3	3.5/2.3	3.2/2.6	4.0/3.1	4.5/2.5	4.2/2.7	3.2/1.9	3.1/2.3	0.9/1.4
11. Storm is a coffeepot	4.0/3.4	3.6/4.0	3.3/2.2	3.4/2.8	3.7/3.0	4.1/2.3	3.9/3.4	3.2/2.1	2.8/2.4	1.2/1.7
12. Smoke is a shroud	4.2/4.7	3.7/5.2	3.5/4.0	3.4/4.1	3.7/3.9	4.1/3.7	3.6/4.7	3.3/3.9	3.0/4.1	0.9/1.4
13. Ant is a bulldozer	4.3/4.9	3.9/4.5	3.7/2.3	3.7/3.6	4.0/3.3	4.4/3.8	4.0/4.2	3.3/3.9	3.2/2.8	0.9/2.1
14. Detective is a ferret	4.3/4.0	3.9/4.0	3.6/3.3	3.2/3.5	3.8/3.3	4.3/3.5	4.0/3.4	3.4/2.7	2.8/3.3	1.0/1.4
15. Lawyer is a lighthouse	4.8/3.8	4.3/3.9	4.1/2.2	4.0/3.2	4.3/2.4	4.4/2.7	4.2/2.7	3.5/2.0	3.3/2.3	1.2/1.3
16. Star is a signpost	4.8/5.4	4.5/5.5	4.1/3.4	4.3/4.1	4.5/3.7	4.7/4.0	4.2/3.6	3.5/3.9	3.5/3.6	1.1/1.8
17. Mosquito is an arrow	4.6/5.0	4.2/5.2	4.0/2.6	3.5/3.7	4.4/3.9	4.7/3.7	4.4/4.0	3.5/2.9	3.3/3.4	1.2/1.7
18. Truth is a firefly	4.6/4.0	3.7/3.9	3.8/2.8	3.8/3.1	4.1/3.2	4.1/2.6	4.3/3.4	3.6/2.5	3.3/3.4	1.1/1.7
19. Dynasty is a play	4.3/3.1	3.9/3.1	3.8/2.7	3.7/3.0	3.9/2.1	3.9/2.7	4.0/2.4	3.6/1.9	3.3/2.4	1.0/1.5
20. Forest is a harp	4.3/3.6	3.7/3.6	3.7/2.4	3.7/2.8	4.1/2.4	4.5/2.7	4.3/2.7	3.6/2.0	3.2/1.9	0.8/1.3
21. Wisdom is a foreigner	4.7/3.6	4.2/3.2	4.1/2.4	4.1/3.0	3.4/2.8	3.7/2.3	3.8/3.1	3.6/1.9	3.4/2.4	1.2/1.5
22. Sailboat is a leaf	4.7/5.2	4.2/5.2	4.0/2.7	3.8/3.4	4.7/3.8	4.7/3.8	4.7/4.4	3.8/2.7	3.5/3.3	1.3/1.6
23. Wish is a rainbow	5.0/4.1	4.7/4.9	4.3/2.3	4.3/3.7	4.7/3.7	4.7/4.3	4.7/4.1	4.0/3.7	3.6/3.6	1.5/1.6

Table 7 (continued)

CMP	ESI	MET	MGD	IMG	IMS	IMP	FAM	SRL	ALT
24. Fireman is a boxer 4.8/3.7	4.6/3.8	4.2/2.3	4.3/2.6	4.6/2.8	5.0/3.2	4.7/2.8	4.0/2.0	3.8/1.9	1.4/1.2
25. Danger is a spice 5.1/4.6	5.0/4.9	4.5/2.9	4.6/3.6	4.8/3.5	4.8/3.7	4.7/3.7	4.0/3.4	3.6/3.5	1.3/1.8
26. Evolution is a lottery 5.2/3.6	4.7/3.1	4.5/2.4	4.4/3.1	4.4/2.1	4.6/3.2	4.6/2.3	4.0/2.5	3.8/2.0	1.3/1.4
27. Dilemma is a cactus 5.1/4.0	4.7/4.4	4.2/2.2	4.1/4.0	4.6/3.2	4.7/3.7	4.6/4.0	4.0/2.4	3.7/3.5	1.2/1.4
28. Artist is a peacock 5.3/4.0	5.0/4.2	4.7/2.7	4.9/2.8	5.1/3.2	5.2/3.0	5.1/3.3	4.1/2.2	4.0/2.7	1.3/1.4
29. Smile is an ambassador 5.0/5.1	4.6/4.4	4.5/2.6	4.7/3.8	4.6/3.4	4.7/4.4	4.5/2.7	4.1/3.4	3.7/3.5	1.3/1.4
30. Time is a physician 4.8/3.7	4.4/4.3	4.2/2.1	4.4/2.7	4.2/2.1	4.6/3.0	4.5/2.4	4.3/3.1	3.9/2.9	1.3/1.5
31. Desert is a sea 5.4/5.3	5.0/4.7	4.6/2.4	4.6/3.7	4.7/4.1	5.0/4.4	4.9/3.7	4.3/4.4	4.0/3.3	1.2/1.8
32. Sleep is an ocean 5.0/3.8	4.8/4.3	4.6/2.5	4.5/2.9	4.8/3.4	4.8/3.6	4.8/2.8	4.4/3.4	3.9/2.4	1.4/1.9
33. Deceit is an ambush 5.0/4.7	4.8/5.0	4.4/3.4	4.3/3.9	4.5/3.2	4.7/3.4	4.5/3.7	4.4/3.0	4.1/3.9	1.1/1.7
34. Butterfly is a flower 5.2/4.8	4.9/5.1	4.5/2.3	4.5/3.3	5.0/4.3	5.4/3.9	5.1/3.9	4.5/3.7	4.2/3.1	1.5/1.6
35. Judge is a balance 5.5/5.4	5.2/5.3	4.9/3.6	4.7/4.9	4.9/5.1	5.0/5.4	5.0/5.0	4.5/3.6	4.3/4.8	1.2/2.0
36. Discipline is a fertilizer 5.3/4.0	5.0/3.8	5.1/2.7	5.0/3.5	5.0/2.9	5.0/3.2	5.0/3.4	4.5/2.3	4.0/2.4	1.5/1.5
37. Blackmailer is a leech 5.9/5.7	5.4/6.1	5.2/3.4	4.9/4.4	5.4/4.3	5.5/4.7	5.2/4.9	4.6/4.7	4.3/5.1	1.5/1.8
38. Alcohol is a crutch 5.5/6.3	5.2/6.4	5.2/4.4	4.9/5.2	5.1/4.9	5.4/4.9	5.2/5.4	4.7/6.6	4.0/5.2	1.4/2.0
39. Indecision is a whirlpool 5.5/5.7	5.5/5.8	5.2/2.9	4.9/5.1	5.3/4.6	5.2/4.7	5.2/5.7	4.8/4.0	4.8/4.8	1.1/2.0
40. History is a mirror 5.4/3.7	5.4/3.6	5.1/2.4	5.0/2.7	5.2/2.1	5.1/3.0	5.2/2.7	5.0/2.1	4.2/2.2	1.6/1.2
41. Loneliness is a desert 6.3/6.2	6.1/6.2	5.8/3.8	5.1/4.8	6.1/5.8	5.6/5.5	5.3/5.8	5.0/4.9	5.1/5.8	1.8/2.2
42. Anger is a blizzard 6.0/5.4	6.0/6.0	5.5/3.2	5.3/4.2	5.9/5.5	5.8/4.7	5.5/5.0	5.0/4.1	4.7/4.1	1.3/2.2
43. Degree is a doorway 5.6/4.6	5.5/3.8	5.3/2.5	5.4/3.6	5.7/3.4	5.6/3.3	5.3/3.4	5.2/4.4	4.9/3.7	1.7/1.5
44. Tumor is a plague 6.1/5.0	6.0/6.1	5.6/3.2	5.1/4.6	5.7/4.2	5.6/4.7	5.3/5.2	5.5/4.6	5.1/4.7	1.3/2.1
45. Tree is an umbrella 6.2/5.6	6.1/5.8	5.8/3.2	6.0/4.9	6.3/4.7	6.3/5.1	6.0/4.2	5.6/4.4	5.3/4.7	1.7/2.0
46. Education is a lantern 5.9/5.8	5.8/6.1	5.8/3.3	5.7/5.3	5.9/4.7	5.9/5.0	5.6/5.1	5.6/4.7	5.2/4.6	1.6/2.0
47. Love is a flower 6.1/5.8	6.1/6.0	5.7/2.9	5.6/3.8	6.1/4.6	5.7/5.0	5.8/4.7	5.6/5.2	5.1/4.3	2.1/3.1

Table 7 (continued)

CMP	ESI	MET	MGD	IMG	IMS	IMP	FAM	SRL	ALT
48. Baby is an angel									
6.4/5.5	6.3/6.1	6.0/5.4	5.8/3.2	6.0/3.9	6.2/5.4	5.7/4.9	5.7/5.5	5.5/4.6	1.9/2.3
49. Gene is a blueprint									
6.3/6.7	6.4/6.6	6.1/4.9	6.0/5.8	6.0/5.6	6.0/5.2	6.0/5.0	5.9/5.9	5.9/5.8	1.5/1.9
50. Mind is a sponge									
6.6/6.3	6.6/6.6	6.4/3.9	6.5/4.9	6.5/5.1	6.4/4.7	6.2/5.4	6.3/5.6	5.9/5.3	1.9/2.2

Averages from the UIC population are listed first, followed by the averages from Katz et al. (1988). Metaphors are listed in order from least to most familiar on the basis of the UIC familiarity score. CMP = comprehensibility, ESI = ease of interpretation, MET = metaphoricity, MGD = metaphor goodness, IMG = metaphor imagery, IMS = subject imagery, IMP = predicate imagery, FAM = familiarity, SRL = semantic relatedness, ALT = number of alternative interpretations.

Appendix C

In the original study performed by Katz et al. (1988), participants rated all of the metaphors on a single dimension. We modified this procedure by having participants rate each metaphor on all ten dimensions. One concern about our method is that the first dimension that was rated (comprehensibility) would impact the ratings of the remaining nine dimensions for a given metaphor. For example, if participants rated the metaphor *The mind is a sponge* as being highly comprehensible, they might continue to rate familiarity, imagery, and other dimensions highly. If this were the case, this might help explain why our interscale correlations were larger than those collected by Katz et al. The purpose of this follow-up study was to address this possibility.

To determine whether the method used for the UIC norms led to higher interscale correlations, we completed a second study replicating the exact method used by Katz et al. (1988). We did this using three dimensions: comprehensibility, familiarity, and number of alternative interpretations. These were chosen because of their places in the order of the dimensions used in our first study. Comprehensibility was the first dimension rated; therefore, it might have influenced the ratings of the other dimensions. The number of interpretations was the last dimension rated; therefore, this dimension would have been most influenced by the ratings of the other dimensions. Familiarity was rated following several other dimensions, but we included it here specifically because familiarity strongly influences metaphor comprehension and plays a key role in several theories of metaphor processing. If we were to find interscale correlations similar to those of Katz et al., we could conclude that the method used in our first study led to inflated interscale correlations. However, if we were to find interscale correlations similar to those we reported in our primary study, we could be confident that the change in methodology did not influence the results.

Methodology

Participants

Ninety undergraduate students from the University of Illinois at Chicago (UIC) participated for credit toward their introductory psychology course. As in the previous study, participants were required to have attended English-speaking schools for at least 10 years in order to ensure English proficiency.

Materials, apparatus, and procedure

The same 50 metaphors used in the primary study were used here. The ratings packets included the 50 metaphors in the same order as in the primary study, but with instructions for rating the metaphors on a single dimension, rather than on all ten as had done in the original study. Participants were assigned to one of three conditions, in which they would rate the metaphors on the basis of comprehensibility, familiarity, or the number of alternative interpretations. After completing the rating portion of the experiment, participants completed the same language history questionnaire and vocabulary test that was used in the primary study.

Results

Overall comparisons

The average ratings for each of the three selected dimensions were collected for each metaphor and then correlated with the original norms collected by Katz et al. (1988), using Pearson correlations. Table 8 presents the average ratings and correlations for the UIC and Katz et al. participants. There were significant, large correlations between the UIC and Katz et al. ratings for each dimension. Note that the correlations found here are very similar to those found in the primary

Table 8 Average ratings (maximum = 7, with standard deviations in parentheses) for each metaphor dimension, correlations between the Katz et al. (1988) and UIC ratings, and difference scores between the Katz et al. and UIC ratings

Dimension	UIC	Katz et al.	Correlation	Difference
Comprehensibility	4.2 (1.2)	4.5 (1.0)	.67***	−0.30*
Felt familiarity	3.7 (1.2)	3.3 (1.3)	.82***	0.42***
Interpretations	1.4 (0.4)	1.6 (0.4)	.62***	−0.25***

Difference scores were calculated by the subtracting the Katz et al. ratings from the UIC ratings, with the absolute difference being reported. Difference scores were compared to zero. Significant correlations and difference scores are indicated by asterisks (* $p < .05$, *** $p < .001$).

study. The largest change in correlations occurred in the comprehensibility dimension, with a $-.05$ shift in correlation.

Interscale correlations

Correlations were computed between the three selected dimensions. Table 9 shows that all of the dimensions are still strongly correlated with one another. The average interscale correlation here (.90) is similar to the average interscale correlation found in our primary study (.93). The largest change in interscale correlation occurred between comprehensibility and number of interpretations, which increased in value by .05 relative to the primary study.

Discussion

The purpose of this follow-up study was to determine whether the methodology used in the primary study led to increased interscale correlations, relative to Katz et al.'s (1988) study. The results support the conclusion that rating the metaphors on one dimension leads to interscale correlations similar to those from rating the metaphors on multiple dimensions. The magnitude of the correlations between the UIC data and Katz et al.'s data was also similar to what we found when participants rated multiple dimensions. These findings indicate

Table 9 Interscale correlations between metaphor dimensions

	CMP	FAM	ALT
CMP	1	.95*	.92*
FAM		1	.82*
ALT			1

All correlations are significant (*) at $p < .001$. CMP = comprehensibility, FAM = familiarity, ALT = number of alternative interpretations.

that the ratings and correlations reported in our primary study are accurate and reliable.

References

- Blasko, D. G., & Briihl, D. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol, 12*, 261–285.
- Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 295–308. doi:10.1037/0278-7393.19.2.295
- Boers, F. (2003). Applied linguistics perspectives on cross-cultural variation in conceptual metaphor. *Metaphor and Symbol, 18*, 231–238.
- Bortfeld, H. (2002). *What native and non-native speakers' images for idioms tell us about figurative language*. Bilingual sentence processing, (pp. 275–295). Amsterdam: North-Holland/Elsevier Science Publishers.
- Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*, 193–216. doi:10.1037/0033-295X.112.1.193
- Campbell, S. J., & Raney, G. E. (2013, November). *The role of familiarity and context strength in metaphor processing*. Paper presented at the annual meeting of the Psychonomic Society, Toronto, Canada.
- Colston, H. L. (2005). *Social and cultural influences on figurative and indirect language*. Figurative language comprehension: Social and cultural influences, (pp. 99–130). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- De Grauwe, S., Swain, A., Holcomb, P., Ditman, T., & Kuperberg, G. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia, 48*, 1965–1984.
- Diaz, M. T., Barrett, K. T., & Hogstrom, L. J. (2011). The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia, 49*, 320–330. doi:10.1016/j.neuropsychologia.2010.12.004
- Diaz, M. T., & Hogstrom, L. J. (2011). The influence of context on hemispheric recruitment during metaphor processing. *Journal of Cognitive Neuroscience, 23*, 3586–3597. doi:10.1162/jocn_a_00053
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language, 37*, 33–355.
- Glucksberg, S. (2008). How metaphors create categories—quickly. In R. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 67–83). Cambridge, UK: Cambridge University Press.
- Kacirik, N. A., & Chiarello, C. (2007). Understanding metaphors: Is the right hemisphere uniquely involved? *Brain and Language, 100*, 188–207. doi:10.1016/j.bandl.2005.10.010
- 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.
- Kuiken, D., Chudleigh, M., & Racher, D. (2010). Bilateral eye movements, attentional flexibility and metaphor comprehension: The substrate of REM dreaming? *Dreaming, 20*, 227–247.
- Littlemore, J., & Low, G. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics, 27*, 268–294.
- Minkoff, S., & Raney, G. (2000). Letter-detection errors in the word *the*: Word frequency versus syntactic structure. *Scientific Study of Reading, 4*, 55–76.
- Schmidt, G., & Seger, C. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition, 71*, 375–386. doi:10.1016/j.bandc.2009.06.001
- Therriault, D. J., & Raney, G. (2007). Processing and representing temporal information in narrative text. *Discourse Processes, 43*, 173–200.
- Thibodeau, P., & Durgin, F. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol, 26*, 206–226.
- Xu, X. (2010). Interpreting metaphorical statements. *Journal of Pragmatics, 42*, 1622–1636.