

Aptness and beauty in metaphor*

MARGUERITE MCQUIRE

Department of Neurology, Center for Cognitive Neuroscience, University of Pennsylvania

LAUREN MCCOLLUM

Department of Neurology, Hospital of the University of Pennsylvania

AND

ANJAN CHATTERJEE

Department of Neurology, Center for Cognitive Neuroscience, University of Pennsylvania

*(Received 20 October 2015 – Revised 23 March 2016 – Accepted 01 May 2016 –
First published online 01 July 2016)*

ABSTRACT

Metaphors are comparisons that link dissimilar conceptual domains. We hypothesized that the aptness of a metaphor is linked to the reader's experience of beauty, and that age and expertise influence these aesthetic judgments. We had young adults, literary experts, and elderly adults rate metaphors for beauty or aptness. Experimental materials consisted of single-sentence novel metaphors whose familiarity, figurativeness, imageability, interpretability, and overall valence ratings were known. Results suggest that beauty and aptness of metaphors are linked for elderly adults but are orthogonal for young adults and literary experts. Elderly participants seem to conflate emotional content with aptness. Young adults are most swayed by a perceived feeling of familiarity when rating for aptness, but not for beauty. Literary experts are relatively unaffected by the psycholinguistic variables, suggesting an emotionally distanced approach to these sentences. Individual differences in literary

[*] This research was supported by a National Institute of Health grant (R01-DC012511) awarded to Anjan Chatterjee, a National Institute of Health training grant (T32AG000255-16). The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Address for correspondence: Marguerite McQuire, Department of Neurology, Center for Cognitive Neuroscience, University of Pennsylvania, 3720 Walnut Street B-51, Philadelphia, PA 19104-6241; tel: (857) 928-3234; e-mail: mcquire@mail.med.upenn.edu

training and life experience have varying effects on the aesthetic experience of metaphor in regard to beauty and aptness.

KEYWORDS: metaphor, aesthetics, aptness, beauty, literary aesthetics.

1. Introduction

Is an apt metaphor also beautiful? Some define the aptness of metaphors as the extent to which concepts are aligned (Chiappe & Kennedy, 1999; Jones & Estes, 2005, 2006). This view of aptness could also relate to beauty. While the aesthetics of visual art, human forms, and natural vistas has been studied extensively (Chatterjee, 2014), we know relatively little about what comprises the aesthetic experience of literary forms (but see Jacobs, 2015, on methods, recent developments, and challenges in neurocognitive poetics). Chatterjee and Vartanian (2014) suggested that aesthetic experiences emerge from an interaction between sensory-motor, emotional valuation, and meaning-knowledge systems. Similarly, Bergen and colleagues have noted that mental imagery or simulations may be used to process literal and figurative language (Bergen, Lindsay, Matlock, & Narayan, 2007; Troyer, Curley, Miller, Saygin, & Bergen, 2014). We chose metaphor for this investigation because, like visual art, its beauty and/or aptness can produce pleasure (Coates, 2002; Crilly, Moultrie, & Clarkson, 2004). However, for metaphors, the meaning-knowledge systems are likely to be more important than the sensory-motor systems.

In order to test the central hypothesis of this study, that apt metaphors are beautiful, we need to deconstruct the notions of aptness and beauty. We wished to identify which, if any, psycholinguistic features contribute to aptness and beauty. To do so, we used novel nominal (i.e., “The X is a Y”) metaphors that we normed extensively (see Cardillo, Schmidt, Kranjec, & Chatterjee, 2010; Cardillo, Watson, & Chatterjee, 2016). Previous research indicates that characteristics of metaphors such as familiarity, figurativeness, imageability, interpretability, and overall valence can influence comprehension of those metaphors (Cardillo et al., 2010; Ianni, Cardillo, McQuire, & Chatterjee, 2014). We used these psycholinguistic features to investigate the underpinnings of aptness and of beauty in metaphors.

The fluency hypothesis in empirical aesthetics states that we prefer stimuli that are easily processed (Reber, Winkielman, & Schwarz, 1998). This hypothesis is most commonly applied to visual objects, but the principle could just as easily apply to literary forms. Aptness in metaphors is typically linked to communicative effectiveness. To understand a nominal metaphor, common characteristics between the source concept and its target are highlighted (Gentner & Wolff, 1997). For example, to understand “His children were his heartbeat”, the reader applies attributes of the source

concept “heartbeat” (e.g., steady, lifelong, source of life) onto the target (children). The resulting “mental space” or blend is new (Coulson, 2001; Fauconnier, 1994). The ease with which this occurs, and the nature of the information learned, could be candidate reasons for a metaphor’s aptness and also for its beauty. In fact, previous researchers have conflated the aptness of a metaphor with its pleasantness. For example, in the only ratings experiment of its kind, Katz, Paivio, Marschark, and Clark (1988) instructed participants to rate a metaphor’s “aptness” by paraphrasing it with “pleasing”. In a later study, Katz expanded the definition of aptness to “pleasing, poetic, surprising” (1989). Preference, or pleasingness, is also used as a measure of beauty.

Familiarity is another candidate characteristic of a metaphor that contributes to its fluency. Just as non-expert viewers of visual art, untrained readers of metaphors seem to “like what they know” (Bohrn, Altmann, Lubrich, Menninghaus, & Jacobs, 2012, 2013; Gerger, 2010). If familiarity is a relevant variable, perhaps literary experts more accustomed to reading metaphors might respond more positively to metaphors than those with less exposure to such literary forms. Additionally, older people, more so than younger people, might have more exposure to figurative language and be more inclined to like metaphors. Similar influences of expertise and world knowledge affect visual aesthetic experiences (Leder, Belke, Oeberst, & Augustin, 2004).

Metaphor research has been plagued by poor stimuli design and a conflation of kinds of metaphors under consideration – nominal and predicate metaphors, familiar and novel metaphors, sentences and phrases (Schmidt, Kranjec, Cardillo, & Chatterjee, 2010). Here, we use relatively novel metaphors because they highlight unexpected relationships. The readers’ initial recognition of the anomaly leads to tension (Reinsch, 1971). The resolution of the tension as the metaphor is understood might produce ‘pleasure’ or ‘relief’ (Sopory & Dillard, 2002). On this view, relatively novel metaphors might be less fluent, but more pleasing because of ‘optimal innovation’ and the possibility for greater incongruity resolution (Giora, 2014; Giora, Fein, Kronrod, Elnatan, Shuval, & Zur, 2004).

Similarly, complexity may play a part in aesthetic experiences. Some findings suggest the elderly prefer stimulus ‘clarity’ and ‘ease’ over complexity, and positive over negative sentiments (Mares, Oliver, & Cantor, 2008). Contrastingly, experts in visual art are more likely than young novices to value complexity and nuance (Bourdieu, 1987). Experts appreciate mild emotional responses, and in general like negative content more than laypersons (Leder, Gerger, Brieber, & Schwarz, 2014).

The present study investigated the effects of sentence-level psycholinguistic characteristics on aptness and beauty ratings made by (a) young adults, (b) literary experts, and (c) elderly adults. These experiments delineate the

extent to which familiarity, imageability, figurativeness, interpretability, and valence contribute to aptness and beauty in people's experience of metaphors. We use the results to test the central hypothesis that apt metaphors are beautiful, and the subsidiary hypotheses that age and expertise influence aesthetic judgments of literary forms.

2. Experiment 1: metaphor aptness

2.1. METHODS

2.1.1. *Materials*

Stimuli for Experiments 1 and 2 consisted of 296 nominal metaphors ("The X is a Y") developed and normed by Cardillo et al. (2010, 2016). Their ratings on multiple sentence-level characteristics, including familiarity, figurativeness, imageability, interpretability, and valence positive ratio, were collected from college-age young adults, and used in the analyses. Ideally, we would have normative data from all three population groups. However, it is common practice for such normative data to be collected in young adults. In brief, familiarity ratings were obtained by asking participants to rate the "frequency of experience with the sentence and its meaning" for each metaphor on a scale from 1 (very unfamiliar) to 7 (very familiar). Figurativeness ratings were obtained by asking participants to rate "how literal of an interpretation each sentence suggested" on a scale from 1 (very literal) to 7 (very figurative). Imageability ratings were obtained by asking participants to rate "how quickly and easily each sentence brought a visual image to mind" on a scale from 1 (no image) to 7 (clear, immediate image). Interpretability ratings were obtained by asking participants to write an interpretation of each sentence. To generate an interpretability score for each item, the number of interpretations deemed plausible by at least two of three independent judges was divided by the total number of interpretations for that item. Valence positive ratio was calculated by asking each participant to categorize the emotional valence of each sentence as "positive valence" or "negative or neutral valence". The resulting percentage of positive valence rating is the "valence positive ratio".

2.1.2. *Procedure*

Participants rated each of the 296 metaphors for aptness on a 7-point Likert scale (1 = low aptness; 7 = high aptness). Aptness was defined as "the extent to which the metaphor's source concept captures important qualities of the metaphor's target concept". Instructions with four examples were provided. Items were presented in random order on a computer screen. Participants were tested individually in a session lasting less than one hour.

2.2. EXPERIMENT 1A: YOUNG ADULTS

2.2.1. *Participants*

Twenty college-age participants were recruited from the University of Pennsylvania community in compliance with procedures established by the university's Institutional Review Board. They were native speakers of English with a mean age of 19.2 years ($SD = 1.2$), fourteen years of education ($SD = 0.9$), twelve females.

2.2.2. *Results*

The mean aptness rating for the 296 nominal metaphors was 4.00 ($SD = 1.68$; min: 1.65; max: 6.25). Aptness correlated positively with familiarity (Pearson $r = 0.741$, $p < .0005$), imageability (Pearson $r = 0.472$, $p < .0005$), and interpretability (Pearson $r = 0.427$, $p < .0005$). Aptness was negatively correlated with figurativeness (Pearson $r = -0.141$, $p = .015$). There was no significant correlation between metaphor aptness and valence positive ratio.

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in aptness ratings. Familiarity, imageability, interpretability, and figurativeness were included in this analysis, as these four parameters correlated significantly with aptness. 58.9% (the adjusted R-squared value) of the variance in metaphor aptness was explained by familiarity ($\beta = 0.667$, $p < .0005$), interpretability ($\beta = 0.135$, $p < .001$), imageability ($\beta = 0.121$, $p < .005$), and figurativeness ($\beta = 0.100$, $p < .013$).

Semi-partial correlation statistics revealed that familiarity was the only predictor variable that made a large **UNIQUE** contribution to the overall variance in aptness. Familiarity accounted for 25.8% of the variance in aptness on its own, while interpretability, imageability, and figurativeness made smaller, though significant, unique contributions to the overall variance in aptness (2.85%, 1.93%, and 1.82%, respectively).

2.2.3. *Summary*

Familiarity was the major variable that predicted aptness ratings made by young adults. The positive correlation suggests familiar metaphors are regarded as highly apt. Aptness also correlated positively with imageability and interpretability. This suggests highly apt metaphors conjure strong visual images and are interpreted more easily.

2.3. EXPERIMENT 1B: LITERARY EXPERTS

2.3.1. *Participants*

Twenty participants were recruited from various higher education institutions. All participants had earned a Master of Fine Arts degree in creative writing,

had published their writings in the last three years, and considered themselves active literary writers. They had subspecialties in Poetry = 8; Fiction = 11; and Creative Non-fiction = 1. They were native speakers of English with a mean age of 33.3 (SD = 6.9), 18.9 years of education (SD = 1.9), eleven females.

2.3.2. *Results*

The mean aptness rating was 3.68 (SD = 0.83; min: 1.67; max: 5.53). Aptness correlated positively with familiarity (Pearson $r = 0.608$, $p < .0005$), interpretability (Pearson $r = 0.407$, $p < .0005$), and imageability (Pearson $r = 0.376$, $p < .0005$). Aptness correlated negatively with figurativeness (Pearson $r = -0.114$, $p < .05$). There was no correlation between metaphor aptness and valence positive ratio.

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in aptness. Familiarity, imageability, interpretability, and figurativeness were included in this analysis, as these four parameters significantly correlated with aptness. The analysis indicated that 40.7% (the adjusted R-squared value) of the variance in metaphor aptness was explained by familiarity ($\beta = 0.532$, $p < .0001$) and interpretability ($\beta = 0.182$, $p < .001$). Figurativeness ($\beta = 0.084$, $p < .013$) and imageability ($\beta = 0.121$, $p < .005$) were not correlated with aptness.

Semi-partial correlation statistics revealed that familiarity was the only predictor variable that made a large UNIQUE contribution to the overall variance in aptness. Familiarity accounted for 17.9% and interpretability accounted for 2.7% of the variance. Imageability, figurativeness, and positive valence ratio did not uniquely contribute to a significant part of the variance in aptness rating.

2.3.3. *Summary*

Like literary novices, experts were more likely to rate familiar than unfamiliar metaphors as apt. As with literary novices, more interpretable metaphors were rated as highly apt. However, imageability and figurativeness did not affect literary experts' judgments of aptness. Experts were not swayed by how easily the metaphor conjured an image when judging the aptness of the metaphor.

2.4. EXPERIMENT 1C: ELDERLY ADULTS

2.4.1. *Participants*

Twenty elderly participants were recruited from the University of Pennsylvania community in compliance with procedures established by the university's

Institutional Review Board. They were native speakers of English, college graduates, with a mean age of 65.3 years ($SD = 6.4$), 17.8 years of education ($SD = 2.7$), thirteen females.

2.4.2. Results

The mean aptness rating was 4.06 ($SD = 0.91$; min: 1.75; max: 6.05). Aptness was positively correlated with familiarity (Pearson $r = 0.706$, $p < .0001$), imageability (Pearson $r = 0.440$, $p < .0001$), interpretability (Pearson $r = 0.429$, $p < .0001$), and valence positive ratio (Pearson $r = 0.183$, $p < .002$). Aptness correlated negatively with figurativeness (Pearson $r = -0.182$, $p < .002$).

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in aptness. Familiarity, imageability, interpretability, valence positive ratio, and figurativeness were included in this analysis, as all five correlated significantly with aptness. 53.7% (the adjusted R-squared value) of the variance in metaphor aptness was explained by familiarity ($\beta = 0.604$, $p < .0001$), interpretability ($\beta = 0.169$, $p < .001$), and valence positive ratio ($\beta = 0.09$, $p < .029$). Imageability ($\beta = 0.085$, $p < .07$) and figurativeness ($\beta = 0.46$, $p < .279$) did not reliably explain variance on aptness rating in elderly adults.

Semi-partial correlation statistics revealed that familiarity uniquely accounted for 23.0%, and interpretability 2.3%, of the variance in aptness ratings. Valence positive ratio positively influenced aptness rating but did not contribute uniquely to variance.

2.4.3. Summary

Elderly adults rated more familiar metaphors as apt, just as young adults and literary experts. Unlike either of the other groups, elderly adults were swayed by the emotional content of the sentence when rating aptness. Metaphors that contained positive words and suggested an overall positive emotional meaning were more likely to be rated as apt by elderly participants.

2.5. ANALYSIS OF APTNESS: THREE GROUPS

Rating data were analyzed using linear mixed effects (LME) models (lme4 package, version 0.999999-2; Bates, Maechler, & Bolker, 2013, in the R Project for Statistical Computing environment, version 3.0.2; R Development Core Team, 2013). LME allows us to investigate variables that are based on subject-related differences (e.g., age and expertise) and item-related differences (e.g., familiarity ratings and positive valence for metaphors).

This kind of analysis cannot be easily accomplished using traditional ANOVA (see Baayen, Davidson, & Bates, 2008, for a more detailed account of the rationale for using LME).

Mean aptness ratings did not significantly differ between the three participant groups [$F(2,57) = 1.51, p = .23$]. Figure 1 shows effect sizes of sentence characteristics for aptness rating in each participant group. The model shows that young adults, literary experts, and elderly adults were similarly influenced by figurativeness, imageability, and interpretability of metaphors. However, the groups diverged in their reliance on familiarity [$F(2,57) = 3.27, p < .045$] and positive valence [$F(2,57) = 7.07, p < .002$] when rating metaphors for aptness. Young and elderly adults' reliance on familiarity ($\beta = 0.58; \beta = 0.51$, respectively) was significantly greater than experts' ($\beta = 0.37, SE = .33, p < .03$). Elderly adults relied significantly more on positive valence of the metaphor ($\beta = 0.28, SE = .14, p < .001$) than young adults ($\beta = -0.25$) or experts ($\beta = -0.29$).

3. Experiment 2: beauty ratings

3.1. METHODS

3.1.1. Procedure

Participants were asked to rate each of the 296 metaphors for beauty on a 7-point Likert scale (1 = not beautiful at all; 7 = very beautiful). Instructions with three examples were provided. Examples were used to clarify the intent and procedure of the experiment. The subjectivity of the ratings was emphasized (e.g., "There is no right answer"). Items were presented in random order on a computer screen. Participants were tested individually in a session lasting less than one hour.

3.2. EXPERIMENT 2A: YOUNG ADULTS

3.2.1. Participants

Twenty college-age participants, who were not enrolled in the aptness study, were recruited from the University of Pennsylvania community in compliance with procedures established by the university's Institutional Review Board. They were native speakers of English with a mean age of 19.2 years ($SD = 1.0$), 14.25 years of education ($SD = 1.0$), fifteen females.

3.2.2. Results

The beauty ratings of the 296 nominal metaphors were analyzed in the same way as in Experiment 1. The mean beauty rating was 3.17 ($SD = 1.55$; min: 1.65; max: 6.25). Beauty was positively correlated with valence positive

Aptness

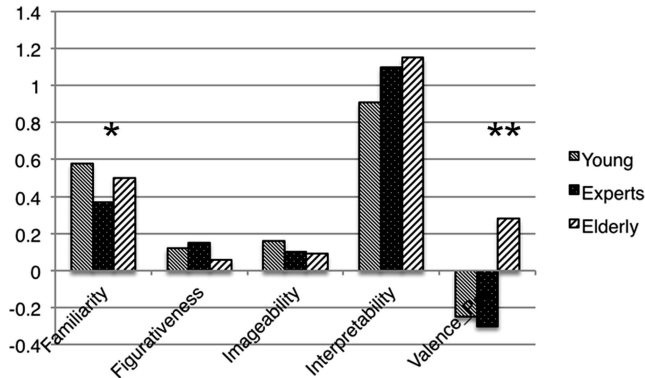


Fig. 1. Effect sizes (*b*) of sentence characteristic for aptness ratings between groups.
 NOTES: * $p < .05$, ** $p < .005$.

ratio (Pearson $r = 0.395$, $p < .0005$), figurativeness (Pearson $r = 0.290$, $p < .0005$), and imageability (Pearson $r = 0.217$, $p < .0005$). There was no significant correlation between beauty of metaphors and their familiarity or interpretability.

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in beauty. Valence positive ratio, figurativeness, and imageability were included in this analysis, as these three parameters were significantly correlated with beauty ratings. The analysis indicated that 27.3% (the adjusted R-squared value) of the variance in metaphor beauty was explained by valence positive ratio ($\beta = 0.390$, $p < .0005$), figurativeness ($\beta = 0.326$, $p < .0005$), and imageability ($\beta = 0.145$, $p < .005$).

Semi-partial correlation statistics revealed that both valence positive ratio and figurativeness made sizeable UNIQUE contributions to the overall variance in beauty ratings. Valence positive ratio accounted for 15.8% of the variance in beauty on its own, while figurativeness accounted for 12.7%. Imageability made a smaller unique contribution (2.59%).

3.2.3. Summary

Valence positive ratio and figurativeness contributed significantly and similarly to beauty ratings made by young adults. The positive relationships between valence expressed and beauty, and between figurativeness and beauty, suggest that young adults without literary expertise associate positive sentiment and more abstract meanings with beauty.

3.3. EXPERIMENT 2B: LITERARY EXPERTS

3.3.1. *Participants*

Twenty participants, who were not enrolled in the aptness study, were recruited from various higher education institutions. All participants had earned a Master of Fine Arts degree in creative writing, were published in the last three years, considered themselves an active literary writer, and had subspecialties in Poetry = 8; Fiction = 8; Creative Non-fiction = 2; and Non-fiction = 2. They were native speakers of English with a mean age of 32.6 (5.9), 19.5 years of education (SD = 1.7), eighteen females.

3.3.2. *Results*

The mean beauty rating was 2.88 (SD = 0.65; min: 1.38; max: 4.63). Beauty correlated positively with figurativeness (Pearson $r = 0.285$, $p < .0001$). Beauty rating correlated negatively with familiarity (Pearson $r = -0.227$, $p < .0001$). There was no significant correlation between beauty rating of a metaphor and imageability, interpretability or valence positive ratio.

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in beauty. Figurativeness and familiarity were included in this analysis as these were significantly correlated with beauty ratings. The analysis indicated that 9.4% (the adjusted R-squared value) of the variance in metaphor beauty was explained by figurativeness ($\beta = 0.235$, $p < .0001$) and familiarity ($\beta = -0.147$, $p < .013$).

Semi-partial correlation statistics revealed that figurativeness and familiarity made small UNIQUE contributions to the overall variance in beauty ratings. These accounted for 4.8% and 1.9% of the variance, respectively.

3.3.3. *Summary*

Literary experts found more figurative and less familiar metaphors to be more beautiful. However, overall the semi-partial correlation suggests that literary experts were not greatly influenced by the psycholinguistic measures previously collected by Cardillo et al. (2010, 2016). Young adults were relatively unaffected by a metaphor's familiarity when rating for beauty. Familiarity had a negative effect on beauty ratings made by literary experts. This suggests that literary experts appreciate novelty and more nuanced links between the source and target of the metaphor.

3.4. EXPERIMENT 2C: ELDERLY ADULTS

3.4.1. *Participants*

Twenty elderly participants, who were not enrolled in the aptness study, were recruited from the University of Pennsylvania community in compliance

with procedures established by the university's Institutional Review Board. They were native speakers of English, college graduates, with a mean age of 65.6 years ($SD = 6.3$), 16.5 years of education ($SD = 3.0$), thirteen females.

3.4.2. Results

The mean beauty rating was 3.64 ($SD = 0.98$; min: 1.60; max: 6.60). Beauty was positively correlated with valence positive ratio (Pearson $r = 0.783$, $p < .0001$), imageability (Pearson $r = 0.294$, $p < .0001$), and familiarity (Pearson $r = 0.154$, $p < .008$). There was no correlation between beauty rating and figurativeness or interpretability.

A multiple regression analysis was performed to determine the extent to which sentence-level characteristics explained variance in beauty. Valence positive ratio, imageability, and figurativeness were included in this analysis, as these three parameters were significantly correlated with beauty ratings. The analysis indicated that 64.0% (the adjusted R-squared value) of the variance in metaphor beauty was explained by valence positive ratio ($\beta = 0.763$, $p < .0001$), imageability ($\beta = 0.135$, $p = .001$), and figurativeness ($\beta = 0.05$, $p < .001$).

Semi-partial correlation statistics revealed that valence positive ratio uniquely accounted for 55.4% of the variance in beauty rating, while imageability and figurativeness accounted for 1.7% and 1.5%, respectively.

3.4.3. Summary

Elderly adults overwhelmingly based their beauty ratings on the emotional content of the metaphor. They exhibited a similar overall pattern as young adults, where familiarity, imageability, and interpretability were positively correlated with beauty ratings.

3.5. ANALYSIS OF BEAUTY: THREE GROUPS

Mean beauty ratings of metaphor differed significantly between the three participant groups [$F(2,57) = 4.14$, $p = .02$]. Figure 2 shows effect sizes of sentence characteristics for beauty ratings in each participant group. The groups diverged in their reliance on familiarity [$F(2,57) = 75.4$, $p < .0001$], figurativeness [$F(2,57) = 10.1$, $p < .0001$], and valence positive ratio [$F(2,57) = 15.6$, $p < .0001$]. Elderly adults relied significantly more on familiarity of the metaphor ($\beta = 0.57$, $SE = .05$, $p < .001$) than young adults ($\beta = 0.02$) or experts ($\beta = -0.16$). However, figurativeness most affected young adults' ratings ($\beta = 0.45$, $SE = .05$, $p < .001$) relative to the elderly

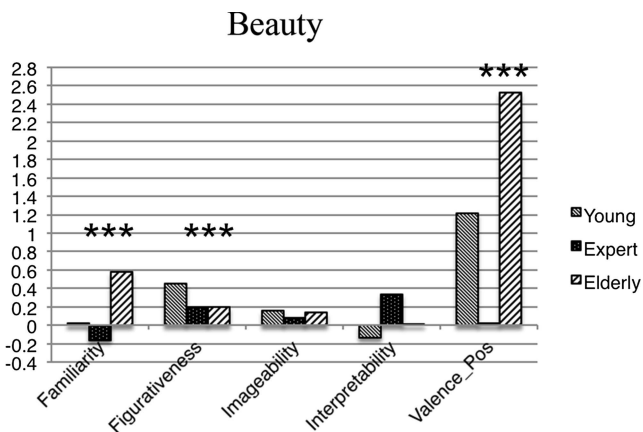


Fig. 2. Effect sizes (β) of sentence characteristic for beauty ratings between groups.
NOTE: *** $p < .0005$.

($\beta = 0.19$) or experts ($\beta = 0.20$). Positive valence of the metaphor heavily influenced elderly adults' beauty ratings ($\beta = 2.53$, $SE = .43$, $p < .001$) relative to young adults ($\beta = 1.25$) or experts ($\beta = 0.02$).

3.6. FURTHER ANALYSIS

We compared beauty and aptness ratings within group (young adults' aptness ratings with young adult's beauty ratings, etc.) and did not find a relationship between beauty and aptness for young adults nor literary expert groups. However, aptness and beauty ratings made by elderly adults correlated significantly [$r(294) = 0.299$, $p < .001$]. Table 1 shows examples of metaphors that were rated highest and lowest for aptness and beauty.

4. General discussion

People use metaphors to express and extend their thoughts. Metaphors capture the essence of ideas that are not communicated easily with literal language. We can marvel at the cleverness with which gifted writers and speakers use figurative language to convey their message. Inherent in such marvel is the fact that we can and often do evaluate language aesthetically. Unfortunately however, we know relatively little about the specific parameters with which we evaluate literary forms. In this study, we examined psycholinguistic characteristics that contribute to two kinds of metaphor valuations – their aptness and their beauty.

The central hypothesis motivating this study is that a metaphor's perceived aptness is linked to its perceived beauty. The subsidiary hypothesis we tested

TABLE 1. *Highest and lowest rated metaphors by participant group*

	High aptness	Low aptness
Young	The coffee was a caffeine bullet.	The wedding planner was a stopwatch.
Experts	The rejection letter was a slap.	The pricey laptop was a cheetah.
Elderly	The opportunity was a career catapult.	The tax refund was a slug.
	High beauty	Low beauty
Young	His children were his heartbeat.	His ugly car is a giggle.
Experts	The love letters were fading footsteps.	The strategy was a media blast.
Elderly	Her eyes were pure laughter.	The tax refund was a slug.

is that expertise and age influences valuations of language. We used well-normed novel metaphors to test these hypotheses (Cardillo et al., 2010, 2016).

The hypothesis that apt metaphors are beautiful was partially confirmed. For elderly participants, aptness and beauty ratings for metaphors were correlated. By contrast, aptness and beauty were orthogonal for young adult and expert groups. Psycholinguistic factors had varying degrees of influence on the aptness and beauty of metaphors for each group. Our second hypothesis was confirmed: age and expertise influence readers' aesthetic experiences of literary entities.

What makes a metaphor apt? In all three groups the feeling of familiarity contributed substantially to aptness judgments. Since our metaphor sentences were novel, this familiarity could not mean that the participants had read these sentences before. Rather, when the idea being conveyed by the metaphor felt familiar, the metaphor felt apt. Memory researchers recognize the distinction between a feeling of familiarity and the recollection of information (Yonelinas, Aly, Wang, & Koen, 2010). An intriguing possibility is that metaphors that make contact with the reader's explicit knowledge are experienced as apt, even when the sentences are encountered for the first time.

Despite the common influence of familiarity on aptness for all three groups, the linear mixed effects model showed granular differences between the groups. Compared to the other groups, experts were less swayed by familiarity when judging metaphors for aptness. By contrast, the elderly, more than other groups, were influenced by positive emotional sentiment in the metaphors. The two groups bring different knowledge and experience to their reading. Perhaps literary experts take a more emotionally distanced, intellectual approach in determining the aptness of metaphors, while the elderly are more likely draw on their real-world emotional experiences.

What makes a metaphor beautiful? Aesthetic experiences emerge out of interaction within an aesthetic triad, between sensory-motor, emotional

valuation, and meaning–knowledge systems (Chatterjee & Vartanian, 2014). While literary forms can evoke sensory-motor memories, the stimuli themselves are impoverished with respect to the immediacy of sensations. As such, one might expect the other parts of the aesthetic triad to have disproportionate influences on the experience of beauty. Our results are consistent with this expectation.

Unlike with judgments of aptness, the groups differed in how beautiful they thought the metaphors were. Literary experts, more than the other groups, were critical of beauty in these sentences. Furthermore, the influence of the psycholinguistic variables on their judgment differed. A notable difference is the effect of positive sentiments expressed in the metaphors. The elderly were influenced a great deal by this variable, young participants to a lesser degree, and experts not at all. Literary experts were also negatively influenced by familiarity, unlike the other two groups. That is, novelty of the idea conveyed in the metaphor contributed to their experience of beauty. Finally, young participants were more affected by figurativeness in judging beauty than the other groups.

Our results show that the fluency hypothesis (Reber et al., 1998) for beauty does not generalize across objects (such as literary forms) and groups of participants. Interpretability, which might be regarded as important for ease of processing, and hence fluency, was not a major factor affecting people's beauty judgments. Familiarity, which might also contribute to ease of processing, was negatively correlated with beauty judgments in literary experts.

In conclusion, we show that apt metaphors are beautiful only for elderly participants. This link is likely mediated by a reliance on positive emotional sentiments for both valuations. In the absence of being grounded by immediate sensory input, compared to visual art for example, individual differences in literary training and life experiences have substantial effects on how people experience beauty in figurative language.

REFERENCES

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
- Bates, D., Maechler, M., & Bolker, B. (2013). lme4: linear mixed-effects modeling using Eigen and syntax. R package [Computer Software]. Vienna: R Foundation for Statistical Computing; online: <<http://CRAN.R-project.org/package=lme4>> (R package version 0.999999-2).
- Bergen, B., Lindsay, S., Matlock, T., & Narayan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*, 733–764.
- Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). Old proverbs in new skins—an fMRI study on defamiliarization. *Frontiers in Psychology*, *3*.
- Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2013). When we like what we know—a parametric fMRI analysis of beauty and familiarity. *Brain and Language*, *124*(1), 1–8.

- Bourdieu, P. (1987). The historical genesis of a pure aesthetic. *Journal of Aesthetics and Art Criticism*, **46**, 201–210.
- 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. *Behavior Research Methods*, **42**(3), 651–664.
- Cardillo, E. R., Watson, C., & Chatterjee, A. (2016). Stimulus needs are a moving target: 240 additional matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, 1–13.
- Chatterjee, A. (2014). Scientific aesthetics: three steps forward. *British Journal of Psychology*, **105**(4), 465–467.
- Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, **18**, 370–375.
- Chiappe, D. L., & Kennedy, J. M. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, **6**(4), 668–676.
- Coates, D. (2002). *Watches tell more than time*. Princeton Junction, NJ: McGraw-Hill Companies.
- Coulson, S. (2001). *Semantic leaps: frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.
- Crilly, N., Moultrie, J., & Clarkson, P. J. (2004). Seeing things: consumer response to the visual domain in product design. *Design Studies*, **25**(6), 547–577.
- Fauconnier, G. (1994). *Mental spaces: aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, **37**(3), 331–355.
- Gerger, G. N. (2010). *Affective and cognitive aspects of aesthetic evaluations in design, art, faces and abstract patterns*. Unpublished doctoral dissertation, Uni Wien.
- Giora, R. (2014). Literal versus nonliteral language: novelty matters. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 330–347). Oxford: Oxford University Press.
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2004). Weapons of mass distraction: optimal innovation and pleasure ratings. *Metaphor and Symbol*, **19**(2), 115–141.
- Ianni, G. R., Cardillo, E. R., McQuire, M., & Chatterjee, A. (2014). Flying under the radar: figurative language impairments in focal lesion patients. *Frontiers in Human Neuroscience*, **8**, 871.
- Jacobs, A. M. (2015). Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception. *Frontiers in Human Neuroscience*, **9**, 186.
- Jones, L. L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, **53**(1), 110–124.
- Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, **55**(1), 18–32.
- Katz, A. (1989). On choosing the vehicles of metaphors: referential concreteness, semantic distances, and individual differences. *Journal of Memory & Language*, **28**(4), 486–499.
- Katz, A., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, **3**(4), 191–214.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, **95**(4), 489–508.
- Leder, H., Gerger, G., Briber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion*, **28**(6), 1137–1147.
- Mares, M. L., Oliver, M. B., & Cantor, J. (2008). Age differences in adults' emotional motivations for exposure to films. *Media Psychology*, **11**(4), 488–511.
- R Development Core Team. (2013). *RA Lang Environ Stat Comput*, **55**, 275–286. Chicago.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, **9**(1), 45–48.
- Reinsch, N. L., Jr. (1971). An investigation of the effects of the metaphor and simile in persuasive discourse. *Speech Monographs*, **38**(2), 142–145.

- Schmidt, G. L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond laterality: a critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society*, **16**(1), 1–5.
- Sopory, P., & Dillard, J. P. (2002). The persuasive effects of metaphor: a meta-analysis. *Human Communication Research*, **28**(3), 382–419.
- Troyer, M., Curley, L. B., Miller, L. E., Saygin, A. P., & Bergen, B. K. (2014). Action verbs are processed differently in metaphorical and literal sentences depending on the semantic match of visual primes. *Frontiers in Human Neuroscience*, **8**, 892.
- Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: examining controversial assumptions and new directions. *Hippocampus*, **20**(11), 1178–1194.