Not All Creative Language Is Created Equal: Subjective Perception and Aesthetic Appreciation of Novel Metaphor and Novel Verbing

Tobias Ungerer (tobias.ungerer@utoronto.ca)[1] & Blair C. Armstrong[1,2]
[1]Department of Psychology, University of Toronto Scarborough
[2]Basque Center on Cognition, Brain and Language, San Sebastián, Spain

Abstract

Experimental findings about linguistic creativity are typically based on a limited set of creative devices, most commonly metaphor. Here, we compare the subjective perception of novel metaphors in English with another prevalent creative device: "verbing," the derivation of new verbs from formally identical nouns (e.g., *to museum the artifacts*). We provide familiarity, meaningfulness, and aesthetic pleasure ratings for a large set of creative verbing and metaphor stimuli with matched controls (n = 486). Both devices were rated as relatively unfamiliar but meaningful. Only novel metaphors, but not verbing, were perceived as highly pleasurable. This partly conflicts with prior work on the relationship between creativity and aesthetic appreciation, which predicts that pleasure should increase for novel but meaningful stimuli. Our findings suggest that the link between creative cognition, as illustrated by different types of novel linguistic devices, and aesthetic appreciation is more complex than previously reported.

Keywords: linguistic creativity; novel verbing; novel metaphor; norming ratings; subjective perception; aesthetic pleasure

Introduction

Language is a primary medium through which creative thoughts—ideas that are both novel and effective (Runco & Jaeger, 2012)—are expressed. Although language users employ numerous semantic-pragmatic and morphosyntactic strategies for creative effect (Jones, 2015), experimental research to date has focused on a small number of these devices while largely disregarding others. In particular, novel metaphors have been the subject of extensive psycho- and neurolinguistic work, yielding valuable insights into the behavioural and neural processes underlying their production and comprehension (for reviews, see Baiocco et al., 2026; Glucksberg, 2003). Some studies explicitly treat novel metaphors as a paradigm case from which conclusions can be drawn about "creative cognition" or "conceptual expansion" at large (e.g., Abraham et al., 2021; Rutter et al., 2012).

However, it remains unclear to what extent findings from novel metaphors generalize to a wider set of creative linguistic devices and thus reflect fundamental aspects of creative cognition. This is especially relevant because few other creative devices have been subjected to a comparable degree of experimental inquiry. A possible exception are novel compounds (e.g., *milk taxi* 'a delivery service dispatched from a farm'), which have received considerable attention (e.g., Bader et al., 2014; Coolen et al., 1993), but these studies have typically not drawn explicit links with linguistic creativity.

In the present study, we examined whether key claims about the subjective perception of novel metaphors generalize to another creative device that has, to date, gained little attention in experimental research on linguistic creativity: the derivation of new verbs from formally identical nouns, known as "noun-to-verb conversion" or simply "verbing" (Clark & Clark, 1979). Verbing is a highly productive word formation process in English (Pinker, 1994) as well as other, typologically unrelated languages such as Mandarin (Cheng & Lasnik, 2016). It occurs both in everyday contexts, such as (1a), and in literary language, such as (1b).

(1) a. After that, we **laptopped** and chatted (https://sandstorm.io/news/2015-05-14-sf-meetup)
b. Oh! To **poet** / like a laser (N. Grimes, *The Dream*)

Existing experimental work on creative verbing, though scarce, has investigated how specific semantic and syntactic factors facilitate the production and comprehension of novel verbs (Gottfurcht, 2008; Kaschak & Glenberg, 2000; Kelly, 1998). Notwithstanding the insights gained from these studies, they have relied on relatively small sets of stimuli that were not normed for linguistic features that play a key role in creativity research, such as familiarity and meaningfulness. A first goal of our study was therefore to provide norming ratings for a large set of novel verbing stimuli and—to enable direct comparisons—novel metaphors along with matched controls, which can scaffold future experimental research on these devices.

The main goal of our study was to test whether specific findings about the perception of novel metaphors generalize to novel verbing. We next turn to a brief description of these findings, before outlining the specific questions of our study.

Prior Evidence and Outstanding Questions About the Perception of Creative Devices

Effects of Familiarity and Meaningfulness A number of studies have investigated the subjective perception of metaphors by collecting ratings on the two core dimensions of familiarity (which distinguishes between conventional and novel metaphors) and meaningfulness (which provides an

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

index of semantic fit), alongside other features such as figurativeness, concreteness, and imageability (for English, see Cardillo et al., 2010; Jankowiak, 2020; Katz et al., 1988; Pissani & de Almeida, 2026; for other languages, e.g., Bambini et al., 2014; Wang, 2022). A subset of these studies compared the relative magnitude of the ratings for novel metaphors with matched controls. Jankowiak (2020) found that novel metaphors (e.g., *viruses are travellers*) were judged as less meaningful than literal controls (e.g., *these people are travellers*) but more meaningful than anomalous controls (e.g., *tables are travellers*; see also Cardillo et al., 2010; Wang, 2022). This intermediate level of meaningfulness relates to evidence from real-time comprehension which shows that novel metaphors are more difficult to process than literal controls but easier than anomalous controls (Arzouan et al., 2007; Obert et al., 2018; Tartter et al., 2002).

Our study investigated whether these trends from metaphor research can also be observed for verbing. This would corroborate the status of this phenomenon as a representative creative (that is, novel but interpretable) device, while also validating our stimulus set for future experimental research on linguistic creativity. Our first question was therefore:

(R1) How do novel verbing and metaphor compare with conventional and anomalous controls in terms of their perceived familiarity and meaningfulness?

Effects of Context Two ratings studies (Bambini et al., 2014; Pissani & de Almeida, 2026) have shown that the perception of metaphors is modulated by the presence of a supportive context, although the specific results varied. Bambini et al. (2014) presented Italian genitive metaphors (e.g., *prato di velluto* 'grass of velvet') in isolation or embedded in multi-sentence literary contexts and found that context led to lower familiarity but higher meaningfulness ratings. In contrast, Pissani and de Almeida (2026) presented English two-word metaphors (e.g., *early bird*) in isolation or embedded in single sentences and found that context led to higher familiarity ratings but did not affect aptness (a measure of semantic appropriateness related to meaningfulness). Either of these findings—that context increases perceived meaningfulness or familiarity—could explain evidence from processing showing that the difficulty in comprehending novel metaphors is alleviated in context (Bambini et al., 2016; Blasko & Briihl, 1997; Hartung et al., 2020).

Extending this line of work, we examined whether context also affects the perception of novel verbing, and whether the effects differ from those for metaphor. For example, it may be the case that a rich global context is more crucial for interpreting metaphors, which involve situationally grounded semantic mappings, than for verbing, which violates morphosyntactic rules on a local structural level.

(R2) How does context modulate perceptions of familiarity and meaningfulness for novel verbing and metaphor?

Aesthetic Appreciation In addition to the abovementioned work, a recent study by Errington et al. (2022) has begun to investigate *aesthetic* perceptions of novel metaphor. In line with the Optimal Innovation Hypothesis (OIH; Giora et al., 2004, 2017), the authors hypothesized that creative metaphors elicit both a "non-default" response (due to their novelty) and a "default" response (by alluding to comprehenders' knowledge of similar conventional metaphors). Errington et al. predicted that the tension between these opposite responses would make novel metaphors more aesthetically pleasurable than both conventional metaphorical controls and anomalous ("excessive") controls. However, this prediction was only partly confirmed: When novel metaphors (e.g., *I brush the meaning*) were presented in isolation, they were rated as *less* pleasurable than conventional controls, contrary to the OIH. However, sentences that contained conventional metaphors followed by novel extensions (e.g., *I grasp the meaning and shake it vigorously*) received the highest pleasure ratings, in line with the OIH. This suggests that aesthetic appreciation may only be maximized when novel metaphors are contextually supported and easy enough to understand.

A final goal of our study was to compare the aesthetic appreciation of novel verbing and metaphor, thus extending recent experimental research on the aesthetics of creative language use (e.g., Errington et al., 2022; Giora et al., 2017). Specifically, we investigated whether, in line with the OIH, the two creative devices were perceived as more pleasurable than controls, and to what extent pleasure was predicted by the interaction of novelty and meaningfulness. In addition, we assessed whether embedding creative expressions in naturalistic contexts—rather than priming them explicitly with the same device (as in Errington et al., 2022)—would enhance aesthetic pleasure. Our third question was:

(R3) To what extent do novel verbing and metaphor give rise to aesthetic pleasure, and how is the latter affected by familiarity and meaningfulness as well as context?

Current Study

Our study was designed to test the three aforementioned research questions. For this purpose, we collected familiarity, meaningfulness, and aesthetic pleasure ratings for instances of novel verbing and novel metaphor, matched with conventional and anomalous controls and presented either in isolation or embedded in a poetic context (as one type of naturalistic environment for linguistic creativity). To our knowledge, our study is the first to directly compare the subjective perception of novel verbing and novel metaphor, using parallel sets of structurally similar stimuli.

Data Availability All stimuli, analysis code, plots, and model outputs are available in the Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/QXMGN

Method

Participants

244 undergraduate students, primarily in their first year of study, were recruited from the Psychology and Linguistics subject pools at the University of Toronto Scarborough. Participants were self-reported native speakers of English, had (corrected-to-)normal vision, and had no history of linguistic, neural, developmental, or motor impairments. They received course credit for participating in the experiment, which lasted approximately 45 minutes.

Materials

Verbing We created 81 novel instances of verbing, such as (2a). Verbs were considered novel if they were unattested as verbs in at least two out of three sources we consulted: the part-of-speech-tagged SUBTLEX-US corpus of movie subtitles (Brysbaert et al., 2012), the WordNet lexical database (Fellbaum, 1998), and the Merriam-Webster Dictionary (n.d.). Each *creative* verbing item was matched with a *conventional* control, such as (2b), in which the critical word functioned as a canonical noun, and an *anomalous* control, such as (2c), in which the critical word was preceded by a short (ungrammatical) adverb or conjunction (max. 4 letters long). The three versions of each triplet were identical except for the preverbal word; given that this word was always short and carried primarily functional meaning, it is unlikely that its lexical characteristics affected the results.

(2) a. We wanted **to museum** the artifacts (creative)
b. We wanted **a museum** to keep the artifacts (conventional)
c. We wanted **very museum** the artifacts (anomalous)

Metaphors In addition, we created 81 novel verb metaphors, such as (3a). Around half of these were based on similar items from a previous norming study (Cardillo et al., 2010). Each *creative* metaphor was matched with a *conventional* control in which the verb was used literally, such as (3b), and an *anomalous* control which contained a noun-verb combination that we considered (largely) uninterpretable, such as (3c).

(3) a. Her **smile clattered** to the floor (creative)
b. Her **keys clattered** to the floor (conventional)
c. Her **beach clattered** to the floor (anomalous)

Using R (R Core Team, 2023), we found no significant differences between conditions at the preverbal nouns in terms of length in letters, lemma frequency (from Brysbaert et al., 2012), orthographic Levenshtein distance (Yarkoni et al., 2008), or concreteness (from Brysbaert et al., 2014; all *p*s > .05), except that metaphors involved less concrete nouns than conventional controls (p = .02). This is an expected property of metaphorical language (Lakoff & Johnson, 1980).

Context Passages For each item triplet, we created a three-line poem that included the target expression in its last line. An example for verbing is shown in (4). The first two lines of the poem, together with the continuation of the third line, are henceforth referred to as "context."

(4) We brushed dust from the attic crates with reverence
Every object a fragile shard of a former life
We wanted to museum the artifacts for our children

Counterbalancing The total set of stimuli (n = 486) was divided into three lists. Each participant rated only one version of each triplet, including both verbing and metaphor, either in isolation or in context. Every item received at least nine ratings for each rating type in each context condition.

AI Usage For a subset of items (fewer than 50%), ChatGPT was used to generate candidates for creative target expressions. In addition, for most items, the AI was used to create a first draft of the poetic contexts. All AI-generated items were thoroughly checked by the first author and a native-speaker consultant. In the majority of cases, substantial modifications were made to increase the naturalness and within-condition consistency of the stimuli.

Procedure

The experiment was conducted online via PCIbex (Zehr & Schwarz, 2018). Each participant rated the items on a 7-point Likert scale along *one* of three dimensions: familiarity (how often they had previously encountered the expression), meaningfulness (how confident they were that they understood the meaning), or aesthetic pleasure (how much they liked or appreciated the way in which the content was expressed). Participants were instructed to use the whole rating scale over the course of the experiment. At the start, they completed several practice trials and received feedback. During the main experiment, participants read each expression, presented either in isolation (for half the participants) or in context (for the other half), and provided their rating. Twenty-five percent of trials were followed by a yes/no comprehension question about the content of either the target expression (presented in isolation) or the first two lines of the poem (presented in context). After the rating phase, participants completed survey-style assessments of their individual (e.g., linguistic and demographic) characteristics, which are not further analyzed here.

Data Exclusions

We excluded 44 participants for whom at least one of the following applied: (a) they answered fewer than 70% of comprehension questions correctly (n = 11); (b) their median rating times were unrealistically short (< 2 s in isolation and < 3 s in context; n = 18); (c) they used fewer than five of the seven scale points throughout the experiment (n = 5); (d) they were among the 5% of participants with the lowest correlation between their ratings and the median by-item ratings of the population (n = 10).

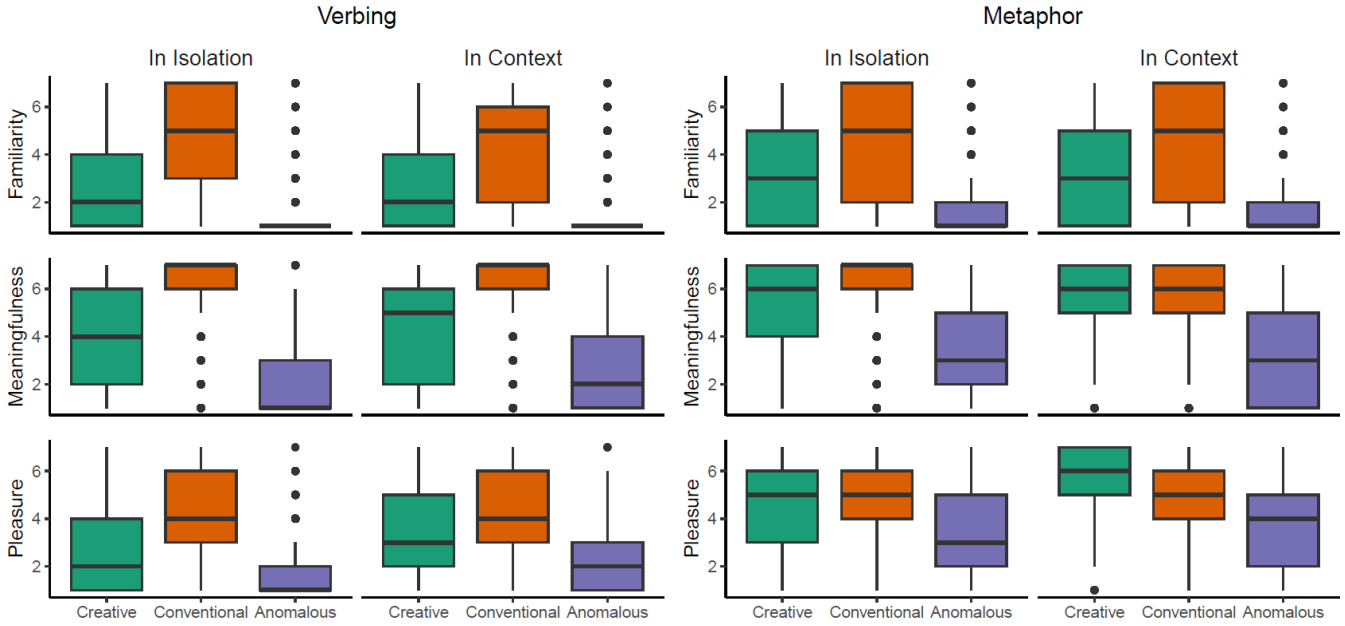

Figure 1: Subjective ratings for verbing, metaphor and their controls, in isolation and in context. Horizontal lines are medians; boxes are interquartile ranges; whiskers extend to maxima within 1.5 × interquartile range.

Results

In the following, we report abbreviated test results; the full regression tables are available in the online repository.

Effects of Creative Devices and Controls, and Impact of Context

We investigated the effects of 'condition' (creative vs. conventional vs. anomalous) and 'context' (in isolation vs. in context) on each rating type (familiarity, meaningfulness, and pleasure). We sum-coded the predictors and fitted separate linear mixed-effects models for each rating type and phenomenon (verbing and metaphor), using *lme4* (Bates et al., 2015). The fixed effects consisted of condition, context, and their interaction. Maximal random effects, consisting of intercepts for participants and items, a by-participant slope for condition, and a by-item slope for context, were included in all models except for the model of pleasure for verbing, where the by-item slope failed to converge. Significance was established via comparison with null models that excluded a given fixed effect using likelihood ratio tests. Main effects and interactions involving condition were further investigated with pairwise post-hoc comparisons between all three levels of condition (at each level of 'context', in the case of interactions), using *emmeans* (Lenth, 2023).

Verbing and its Controls The descriptive results are summarized on the left of Figure 1. There were main effects of condition for all three rating types, familiarity ($\chi^2(2) = 127.9, p < .001$), meaningfulness ($\chi^2(2) = 131.5, p < .001$) and aesthetic pleasure ($\chi^2(2) = 91.4$, $p < .001$). Creative verbing was rated as less familiar, less meaningful, and less pleasurable than conventional controls and more familiar, more meaningful, and more pleasurable than anomalous controls (all ps $< .001$). There were no significant main effects of or interactions with context (all ps $> .05$).

Metaphor and its Controls The ratings are summarized on the right of Figure 1. There were main effects of condition for familiarity ($\chi^2(2) = 135.9$, $p < .001$) and meaningfulness ($\chi^2(2) = 121.8, p < .001$), with creative metaphors being rated as less familiar and less meaningful than conventional controls but more familiar and more meaningful than anomalous controls (all ps $< .001$). Pleasure showed an interaction between condition and context ($\chi^2(2) = 8.2$, $p = .02$). Post-hoc comparisons indicated that, in isolation, metaphors were rated as more pleasurable than anomalous controls ($p < .001$) but did not differ from conventional controls ($p = .79$); whereas in context, metaphors were rated as more pleasurable than both conventional controls and anomalous controls (both ps $< .001$). These differences were driven by the creative condition, which was rated as more pleasurable in context than in isolation ($p = .003$).

Effects of Familiarity and Meaningfulness on Aesthetic Pleasure

In a second step, we examined the relations between rating types by testing how familiarity and meaningfulness predicted pleasure. Since the relationship between these continuous variables should hold across the entire set of items, regardless of condition, we collapsed our analysis across all creative, conventional, and anomalous stimuli.

We calculated by-item means for each rating type and then fitted separate linear regression models for each phenomenon and each context condition. All continuous variables were mean-centered for the analyses (but are plotted below on their

original scales). Significance was evaluated with two-sided *t* tests on regression coefficients. In case of significant interactions, we conducted simple slope analyses with *emmeans*, as detailed below.

Verbing and its Controls When the stimuli were presented in isolation, there was a main effect of meaningfulness, indicating that more meaningful items were perceived as more pleasurable ($p < .001$). There was no main effect of familiarity nor an interaction between the two predictors (both $ps > .05$). These results are illustrated in the upper left panel of Figure 2, which shows how pleasure changed as a function of familiarity at low (–1 SD), mean, and high (+1 SD) levels of meaningfulness. The nearly horizontal regression lines suggest that familiarity had no effect on pleasure, whereas the vertical differences between the lines indicate that higher meaningfulness led to more pleasure.

In context, however, we found an interaction between familiarity and meaningfulness ($p < .001$), suggesting that as one variable increased, the effect of the other variable on pleasure was modulated. As the upper right panel of Figure 2 illustrates, the more meaningful items were, the more attenuated the effect of familiarity on pleasure was. Nevertheless, a simple slopes analysis indicated that familiarity had a positive effect on pleasure at all three levels of meaningfulness (all $ps < .01$).

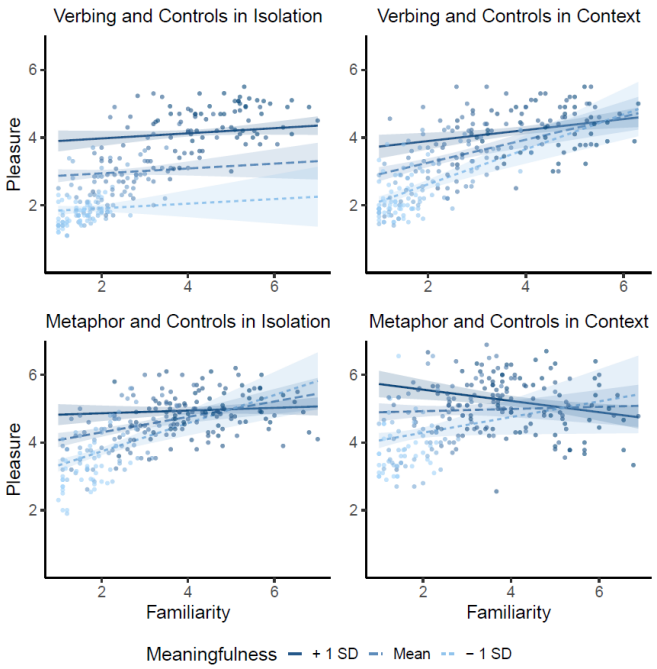

Figure 2: Aesthetic pleasure for verbing and metaphor and their controls as a function of familiarity at three levels of meaningfulness, in isolation or in context. Dots represent mean item-level ratings.

Metaphor and its Controls We found interactions between familiarity and meaningfulness both in isolation and in context (both $ps < .001$). This again indicates that the effect of familiarity on pleasure became weaker as meaningfulness increased. Simple slope analyses confirmed that in isolation (see the lower left panel of Figure 2), familiarity had a positive effect on pleasure at low and mean levels meaningfulness (both $ps < .001$) but no effect at high levels of meaningfulness ($p = .37$). In context (see the lower right panel of Figure 2), the effect of familiarity was fully reversed at different levels of meaningfulness. Higher familiarity gave rise to increased pleasure when meaningfulness was low ($p = .04$), had no effect on pleasure when meaningfulness was at mean ($p = .65$), and led to *reduced* pleasure when meaningfulness was high ($p = .003$).

Discussion

The results of our rating study inform each of the three research questions described above. Below, we first address how familiarity and meaningfulness ratings compared across novel verbing and novel metaphor (RQ1), and to what extent they were modulated by context (RQ2). We then discuss the results for aesthetic appreciation and its relationship with the other rating types as well as context (RQ3).

Familiarity and Meaningfulness Ratings Across Creative Devices

The patterns of results for familiarity and meaningfulness were relatively similar across novel verbing and novel metaphor. For each device, the creative items were rated as less familiar and less meaningful than conventional controls but more familiar and more meaningful than anomalous controls. In terms of absolute magnitudes, both creative devices were rated as rather unfamiliar (by-condition medians of 3 or lower) but relatively meaningful (medians of 4 or higher; see Figure 1). This aligns with the definition of creativity as the combination of novelty and effectiveness (Runco & Jaeger, 2012) and corroborates the internal validity of our three-way distinction between creative, conventional, and anomalous stimuli.

Context did not modulate familiarity or meaningfulness for either verbing or metaphor. No statistically significant differences emerged depending on whether the stimuli were presented in isolation or embedded in a three-line poem. This contrasts with prior studies of metaphor, which have found effects of context on several rating dimensions (Bambini et al., 2014; Pissani & de Almeida, 2026). However, as we noted earlier, the specific effects varied between studies: Whereas Bambini et al. (2014) found that context *decreased* perceived familiarity and increased meaningfulness, Pissani and de Almeida (2026) found that context *increased* familiarity ratings and had no effect on aptness (which is related to meaningfulness). These differences may partly be due to design factors, given that Bambini et al. (2014) used longer contexts and tested only novel metaphors, whereas Pissani and de Almeida (2026) also included conventional metaphors. Still, given the variation among previously reported results, the absence of context effects on familiarity and meaningfulness in our study may be less surprising than it initially seemed. Follow-up work could establish whether

further manipulations of our item contexts (e.g., in length or plausibility) may have more substantial effects on these rating dimensions (but see also below for a discussion of context effects on aesthetic pleasure ratings).

Aesthetic Appreciation and Optimal Innovation

Our results also shed light on a third and less frequently studied dimension of subjective perception: the aesthetic pleasure invoked by creative expressions. When comparing creative conditions with their respective controls, intriguing differences emerged between the two creative devices. Novel verbing was perceived as less pleasurable than conventional controls, although it was more pleasurable than anomalous controls. In contrast, novel metaphors were rated as equally pleasurable as conventional controls when presented in isolation and as more pleasurable than conventional controls when presented in context.

These results only partly support the Optimal Innovation Hypothesis (OIH; Giora et al., 2004, 2017), which predicts optimally innovative stimuli to be perceived as highly pleasurable. Both of our creative devices meet, in principle, the criteria for optimal innovation by eliciting a novel "non-default" response (as illustrated by the low familiarity ratings) but being nevertheless recoverable through their association with a salient "default" response (as suggested by the high meaningfulness ratings). Our results for novel metaphor, at least when presented in context, align with the OIH. The fact that pleasure was higher in context than in isolation can be attributed to the fact that metaphors typically naturally occur, and are appreciated, in rich linguistic environments, such as the poetic contexts used here. This also aligns with Errington et al.'s (2022) results, who found that novel metaphors were only highly pleasurable when the preceding sentence context contained a related conventional metaphor. Our results demonstrate that such explicit contextual priming with the same semantic device is not the only way in which high pleasure can be induced; rather, the presence of a more general poetic context alone can be sufficient for novel metaphors to be perceived as pleasurable, potentially optimal innovations.

Our results for novel verbing, on the other hand, do not support the OIH. This suggests that the OIH may not universally apply across creative devices, and that a more nuanced view of the relationship between different types of linguistic innovations and aesthetic appreciation is called for. As such, prior studies on metaphor alone may not be a good basis for drawing broad, general conclusions about creative cognition.

There are several possible explanations for why verbing, even in context, did not elicit high aesthetic pleasure. First, verbing is primarily creative on a morphosyntactic level, even though it also poses the semantic challenge of deriving an 'activity' meaning from the 'entity' meaning of its corresponding noun. Nevertheless, verbing may not engage language users' attempts at conceptualization and meaning-making to the same extent as metaphors, which entail new mappings between distinct conceptual domains (Lakoff & Johnson, 1980). As a result, verbing may less closely connected to the experiential basis of aesthetic appreciation. Second, whereas metaphor is a prototypical feature of poetic language, verbing may not be typically associated with literary genres (although it is attested in poetry, as illustrated by (1b) above). In particular, many of our verbing stimuli related to mundane concepts, such as *museum* in (2a), which may have made them appear less natural and therefore less enjoyable in poetic contexts. This suggests that, even though different devices may be equally creative, some of them may be better suited for literary use and aesthetic exploitation than others. Follow-up research could investigate whether creative verbing elicits more (aesthetic) pleasure if embedded in non-literary genres (e.g., journalistic texts or casual conversation).

An additional factor revealed by our analysis was that item-level differences modulated the impact of novelty and meaningfulness (as the two core components of creativity) on aesthetic appreciation. Specifically, for both devices and their controls (albeit only in context for verbing), we found interaction effects of familiarity and meaningfulness, indicating that as expressions were perceived as more meaningful, the positive effects of familiarity on pleasure were attenuated or even reversed. This means that highly meaningful items, especially when presented in context, may be more pleasurable if they are novel rather than familiar. These results provide evidence for the OIH but also suggest that the relationships it predicts may only hold for creative expressions that are perceived as highly meaningful.

Conclusion

Experimental research on linguistic creativity to date has been skewed towards investigating certain creative devices, most notably novel metaphors, while leaving others understudied. Here, we have expanded this work by investigating the subjective perception of novel verbing, a highly prevalent creative device in languages such as English.

Our results demonstrate that creative verbing and novel metaphors share aspects of their subjective perception but also differ in important respects, especially regarding their aesthetic pleasure effects. Metaphors were not only perceived as more aesthetically pleasing than verbing, but their degree of aesthetic appreciation was also more malleable to context effects. This illustrates that not all creative devices are created equal, and that researchers should be careful not to draw conclusions from one specific device about linguistic creativity or creative cognition at large.

Moreover, our findings raise questions about the relationship between creativity and aesthetic pleasure, which may be closely related in some cases (e.g., for novel metaphors in our study) but partly dissociated in others (e.g., for our verbing instances). Further in-depth studies across a range of creative phenomena will be needed to unravel these relationships. We hope that our extensive set of matched and normed stimuli will serve as an open resource for a collective effort to further deepen our understanding of linguistic creativity and its links with subjective perception and aesthetic appreciation.

Acknowledgments

We thank Rebecca Norman for assistance with the stimulus creation. This work was supported by the Social Sciences and Humanities Research Council (SSHRC) under Grant No. IDG 522610 and by the Office of the Vice-Principal Academic & Dean at the University of Toronto Scarborough.

References

Abraham, A., Rutter, B., & Hermann, C. (2021). Conceptual expansion via novel metaphor processing: An ERP replication and extension study examining individual differences in creativity. *Brain and Language*, *221*, 105007. https://doi.org/10.1016/j.bandl.2021.105007

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *1160*, 69–81. https://doi.org/10.1016/j.brainres.2007.05.034

Bader, R., Opitz, B., Reith, W., & Mecklinger, A. (2014). Is a novel conceptual unit more than the sum of its parts?: FMRI evidence from an associative recognition memory study. *Neuropsychologia*, *61*, 123–134. https://doi.org/10.1016/j.neuropsychologia.2014.06.006

Baiocco, L., Kielar, A., & Lai, V. T. (2026). Metaphor comprehension in neurotypical adults: A scoping review of event-related potential studies. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-025-01376-z

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00559

Bambini, V., Resta, D., & Grimaldi, M. (2014). A dataset of metaphors from the italian literature: Exploring psycholinguistic variables and the role of context. *PLOS ONE*, *9*(9), e105634. https://doi.org/10.1371/journal.pone.0105634

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

Blasko, D. G., & Briihl, D. S. (1997). Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol*, *12*(4), 261–285. https://doi.org/10.1207/s15327868ms1204_4

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

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. https://doi.org/10.3758/BRM.42.3.651

Cheng, J., & Lasnik, H. (2016). Parametric variations in Mandarin and English denominal verb derivation. *Lingua*, *180*, 25–48. https://doi.org/10.1016/j.lingua.2016.03.006

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, *55*(4), 767–811. https://doi.org/10.2307/412745

Coolen, R., Van Jaarsveld, H. J., & Schreuder, R. (1993). Processing novel compounds: Evidence for interactive meaning activation of ambiguous nouns. *Memory & Cognition*, *21*(2), 235–246. https://doi.org/10.3758/BF03202736

Errington, P. J., Thye, M., & Mirman, D. (2022). Difficulty and pleasure in the comprehension of verb-based metaphor sentences: A behavioral study. *PLOS ONE*, *17*(2), e0263781. https://doi.org/10.1371/journal.pone.0263781

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. MIT 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. https://doi.org/10.1207/s15327868ms1902_2

Giora, R., Givoni, S., Heruti, V., & Fein, O. (2017). The role of defaultness in affecting pleasure: The Optimal Innovation Hypothesis revisited. *Metaphor and Symbol*, *32*(1), 1–18. https://doi.org/10.1080/10926488.2017.1272934

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. https://doi.org/10.1016/S1364-6613(02)00040-2

Gottfurcht, C. A. (2008). *Denominal verb formation in English* [Doctoral dissertation, Northwestern University]. ProQuest. https://www.proquest.com/openview/150d1ed1416a0d53fe194190f2554a2b

Hartung, F., Kenett, Y. N., Cardillo, E. R., Humphries, S., Klooster, N., & Chatterjee, A. (2020). Context matters: Novel metaphors in supportive and non-supportive contexts. *NeuroImage*, *212*, 116645. https://doi.org/10.1016/j.neuroimage.2020.116645

Jankowiak, K. (2020). Normative data for novel nominal metaphors, novel similes, literal, and anomalous utterances in Polish and English. *Journal of Psycholinguistic Research*, *49*(4), 541–569. https://doi.org/10.1007/s10936-020-09695-7

Jones, R. (Ed.). (2015). *The Routledge handbook of language and creativity*. Routledge. https://doi.org/10.4324/9781315694566

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, *43*(3), 508–529. https://doi.org/10.1006/jmla.2000.2705

Katz, A. N., 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. https://doi.org/10.1207/s15327868ms0304_1

Kelly, M. H. (1998). Rule- and idiosyncratically derived denominal verbs: Effects on language production and comprehension. *Memory & Cognition*, *26*(2), 369–381. https://doi.org/10.3758/BF03201147

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

Lenth, R. (2023). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.5*. https://CRAN.R-project.org/package=emmeans

Merriam-Webster. (n.d.). *The Merriam-Webster Online Dictionary*. Retrieved https://www.merriam-webster.com

Obert, A., Gierski, F., & Caillies, S. (2018). He catapulted his words from the dais: An ERP investigation of novel verbal metaphors. *Journal of Neurolinguistics*, *47*, 59–70. https://doi.org/10.1016/j.jneuroling.2018.02.008

Pinker, S. (1994). *The language instinct*. William Morrow.

Pissani, L., & de Almeida, R. G. (2026). Metaphors in context and in isolation: Familiarity, aptness, concreteness, metaphoricity, and structure norms for 300 two-word expressions. *Behavior Research Methods*, *58*(1), 31. https://doi.org/10.3758/s13428-025-02902-0

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Rutter, B., Kröger, S., Stark, R., Schweckendiek, J., Windmann, S., Hermann, C., & Abraham, A. (2012). Can clouds dance? Neural correlates of passive conceptual expansion using a metaphor processing task: Implications for creative cognition. *Brain and Cognition*, *78*(2), 114–122. https://doi.org/10.1016/j.bandc.2011.11.002

Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain and Language*, *80*(3), 488–509. https://doi.org/10.1006/brln.2001.2610

Wang, X. (2022). Normed dataset for novel metaphors, novel similes, literal and anomalous sentences in Chinese. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.922722

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Zehr, J., & Schwarz, F. (2018). *PennController for internet based experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832