Liquid Gold Down the Drain: Measuring Perceptions of Creativity Associated with Figurative Language and Play

Stephen Skalicky
School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand
stephen.skalicky@vuw.ac.nz

Abstract

The purpose of this study is to examine layperson perceptions of creativity associated with figurative language and language play. To do so, participants wrote attention-grabbing responses for two news stories and rated whether their responses were less, equally, or more creative when compared to preconstructed responses containing different combinations of metaphor and sarcasm. Participants’ answers were also analyzed for the presence of figurative language or language play. Results demonstrated participants were less likely to self-rate their answers as more creative when compared to preconstructed responses containing figurative language, but only for specific instances of metaphor and sarcasm. Moreover, participants who included figurative language or language play in their responses were significantly more likely to self-rate their answers as more creative. These results suggest layperson perceptions of creativity are influenced by figurative language and language play in a manner which supports scholarly understandings of the relationship between language and creativity.

Keywords

metaphor – sarcasm – figurative language – language play – creativity
1 Introduction

It seems natural to associate creativity with figurative language. Indeed, Aristotle claimed metaphor is wielded by the most intelligent and wittiest of writers (Aristotle, 350 B.C.E./2008). Writing advice such as this has contributed to metaphor and other types of figurative language being described as an ornamental flourish which deviates from standard, literal language use (Gibbs, 1994, 2018; Glucksberg, 2001). But as research has shown, figurative language is not just the application of a different font to the same sentence. Rather, instances of metaphor, verbal irony, and other types of figurative language serve specific and unique communicative and pragmatic functions (Gibbs & Colston, 2012; Veale, 2012). From this point of view, figurative language is subject to the same creative restraints as any other type of language, literal or otherwise, and is thus not inherently creative or restricted to people with greater creative ability (Carter, 2015; Vásquez, 2019).

However, there are still strong links between creativity and figurative language. Even if figurative language is not inherently more creative, it might yet possess greater potential to be used creatively (Gerrig & Gibbs, 1988), which in turn might amplify perceptions of creativity associated with figurative language use. Indeed, some studies have described a positive association between cognitive measures associated with creative ability and ratings of creativity in figurative language (Beaty & Silvia, 2013; Huang et al., 2015; Silvia & Beaty, 2012, 2021). Moreover, purposeful uses of figurative language can be considered a form of language play, which involves intentional manipulation of the form and/or meaning of language for enjoyment or pleasure (Bell, 2016; Cook, 2000). As such, there may still be a privileged role for figurative language when exploring creativity as it relates to language, but this role is far from straightforward. The goal of this study is to further explore this role by examining perceptions of creativity associated with figurative language use.

2 Creativity and Language

One of the primary challenges associated with any discussion of language and creativity is the definition of creativity itself. Linguistically, all language is inherently creative because language is a system which allows for limitless individual variation within the structural constraints of a grammar. Languages are further creative because they allow for the emergence of new constructions, meanings, and uses (Katz & Hussey, 2011). However, this view of the inherent creative potential of all language does little to resolve the question of whether figurative language is more creative than other types of language.
2.1 Creativity and Figurative Language

Indeed, empirical research in computational and corpus linguistics attests to the fluidity of figurative language in terms of its adherence to the constraints of linguistic structure. On the one hand, there are some instances of figurative language which are quite fixed and arguably reliant on their linguistic structure, such as ironic similes (Veale, 2012) and instances of negation which produce, by default, metaphorical or ironic meaning (Giora et al., 2013). Conversely, there is a great deal of linguistic evidence to suggest figurative language and humor may result when a speaker purposefully deviates from linguistic conventions, such as using infrequent collocations or violating the semantic preference of lexical items. For instance, a reversal of typical evaluations associated with entities or actions has been associated with verbal irony (Partington, 2007, 2011). Additionally, situations which prompt a forced reinterpretation of synonymous meanings (e.g., The Drinking Problem Hypothesis, Hoey, 2005) can result from a purposeful use of lexical combinations which are infrequent or otherwise unexpected (Goatly, 2012, 2017; Skalicky, 2018). In these circumstances, purposeful play with expected patterns and uses of language have been argued to fuel perceptions of humor and creativity in language (Hoey, 2005). Taken as a whole, this research suggests that linguistic form alone is only part of the puzzle when considering creative uses of figurative language.

Aside from the differences in use of expected or unexpected linguistic structures, it may still be the case that some instances of figurative language are more creative than others. This point is explicated in Gerrig and Gibbs (1988), who provided a definition of creative language scoped to more specific uses, in which creative language is “any utterance, phrase, or word whose meaning varies with the context in which it is produced in a way that could not be predicted from the lexicalized meaning of its component words” (1988: 2). Although this perspective recognizes any example of language could be used creatively, the definition of creativity is similar to definitions of figurative language: “speech where speakers mean something other than what they literally say” (Gibbs & Colston, 2012: 1). The overlap between these two definitions suggests figurative language has an increased potential for being considered creative because, by definition, figurative meaning is more than the semantic decomposition of the literal form.

2.2 Creativity, Metaphor, and Sarcasm

Two of the most widely studied types of figurative language are metaphor and verbal irony (Gibbs & Colston, 2012). The two differ in a number of ways. First, in terms of structure, metaphor (and similes) are figurative, conceptual associations made in order to explain or make sense of things (Gibbs, 2011; Lakoff
and these associations may also reflect ideological leanings towards entities depicted in a metaphor (Charteris-Black, 2004). Verbal irony is a figurative strategy for providing a statement or opinion which somehow clashes or is inappropriate within the pragmatic and situational context in which it is spoken (Attardo, 2000; Colston, 2017). Secondly, cognitive strategies for understanding metaphorical meaning are thought to differ from verbal irony because verbal irony relies on metarepresentational inference of speakers’ attitudes and thoughts, whereas metaphor relies on knowledge of conceptual domains (Colston & Gibbs, 2002). Along the same lines, cognitive individual differences have been shown to influence metaphor and verbal irony processing in different ways (e.g., Olkoniemi et al., 2016), which further suggests a qualitative difference in how metaphor and verbal irony are processed and comprehended.

A series of elicited figurative language production experiments suggest a third difference between metaphor and verbal irony in terms of their connections to creative ability (Beaty & Silvia, 2013; Huang et al., 2015; Silvia & Beaty, 2012, 2021; Skalicky, 2020). In these studies, creativity is typically defined as a cognitive ability following a definition from psychology which identifies two main components of creativity: originality and effectiveness (Runco & Jaeger, 2012). For metaphor specifically, results from these studies cohere to suggest that cognitive measures associated with creative ability, such as different measures of intelligence, are associated with variance in perceptions of creativity related to metaphor production. However, these effects further depend upon whether an individual is constrained in their linguistic choices when making a metaphor (e.g., being asked to craft a metaphor describing professors as smart), or whether they are relatively free to craft a metaphor of their choosing (e.g., being asked to describe the most disgusting thing one has ever eaten). For instance, participants with higher fluid intelligence scores or who exhibited a greater desire to engage in cognitively difficult tasks produced metaphors rated as more creative, but only when creating unconstrained, novel metaphors (i.e., participants were less able to draw from metaphors they had heard before; Beaty & Silvia, 2013; Silvia & Beaty, 2012; Skalicky, 2020). Similar attempts have been made to link cognitive ability to verbal irony and sarcasm use. One study found that individuals who recalled, produced, or listened to sarcasm reported increased levels of abstract thinking, which in turn boosted

---

1 Although there are a number of different ways to use verbal irony, the focus of the current study is on sarcasm, a specific type of verbal irony used to express negative evaluation from the perspective of the speaker (Colston, 2017). The references to verbal irony in this section thus also incorporate sarcasm.
performance on subsequent tests of creative ability (Huang et al., 2015). However, no relationship was found between sarcasm and abstract thinking in a different study where participants were explicitly asked to be sarcastic (Skalicky, 2020).

3 Current Study

It follows then that figurative language should not be equated with creative language. However, there is theoretical and empirical evidence to suggest that figurative language may possess an increased potential to be defined as creative language, and different types of figurative language such as metaphor and sarcasm may diverge in their relation to other creative processes, such as cognitive measures associated with creative ability. While differences in perceptions of creativity have been reported for both metaphor and sarcasm, these studies typically employ (for good reason) strict operationalizations of creativity based on theoretical definitions. However, because metaphor and sarcasm pervade everyday language use, it would be fruitful to test whether the presence or absence of metaphor and/or sarcasm influence behavior and perceptions towards creativity. At the same time, measuring whether language users employ metaphor, sarcasm, or some other form of figurative language or play when explicitly asked to be creative can provide a better understanding into layperson perceptions of creativity and language. Accordingly, the purpose of this study is to further probe the relationship between creativity and figurative language using these everyday perceptions. The following research questions guide the current study:

1. For everyday language users, are perceptions of creativity towards language influenced by the presence of figurative language?
   1.1. If so, are there differences in these perceptions between different types of figurative language, such as sarcasm and metaphor?

2. When asked to be creative, will participants include figurative language or play in their responses?
   2.1. If so, does the inclusion of figurative language or play in their response influence their self-ratings of creativity?

4 Method

The purpose of the current study is to further investigate the relationship between language and creativity using human perceptions of creativity. Rather
than asking trained raters or experts to assess the creativity of metaphor or sarcasm using rubrics or other standardized operationalizations (e.g., Silvia & Beaty, 2021; Skalicky, 2020), the approach taken in the current study is to examine self-ratings of creativity when compared to language containing metaphor and/or sarcasm. To do so, perceptions of creativity were gathered from participants taking part in a creative response production task. In the task, participants rated the creativity of their own answers when compared to pre-constructed answers containing no figurative language, metaphor, or sarcasm. Crucially, these self-judgments were made without the explicit knowledge that some answers did or did not contain figurative language, which allows for insight into the automatic, holistic response to an answer from a layperson perspective as opposed to calculated analysis. Moreover, statistical analysis comparing the self-ratings of creativity made in response to answers not containing metaphor or sarcasm can determine whether variations in self-rating behavior are due to chance or differences in the response conditions (RQ1).

In this study, the self-ratings of creativity reflect an implicit evaluation of the preconstructed comparison answers (and by extension, the figurative language contained within them). However, although self-ratings of creativity can provide accurate assessments of creativity (Silvia et al., 2012), self-ratings can be also be confounded by variables such as mood (Montgomery et al., 2004) or personality (Reiter-Palmon et al., 2012). Therefore, participant answers were also assessed for the presence or absence of metaphor and sarcasm and other overt attempts at language play, allowing for an alternative measure of whether perceptions of creativity were associated with figurative language (RQ2). Statistical analyses were then carried out to determine whether there was a statistically significant association between participant self-ratings of creativity, the different types of figurative language in the pre-constructed responses, and inclusion of figurative language or play in the participant responses.

4.1 Task Material
Two news stories were selected from the American Voices section of the satirical news website The Onion in order to elicit creative responses from participants. This section differs from the original satirical news stories published by The Onion in that American Voices presents a headline and brief summary of a real news story alongside fictional, satirical responses to the news story purported to be from regular readers of the website (but who are in fact fictional personas). Two stories from this section which contained no overt political or religious content were selected in order to avoid emotional and attitudinal reactions which might bias the way information in the stories was perceived.
Four fictional responses were then written for each of the two stories. These preconstructed responses were designed to contain different types of figurative language while still expressing approximately the same intended meaning. For each story, these preconstructed responses comprised a baseline response, a primarily sarcastic response, a primarily metaphorical response, and a combined sarcastic and metaphorical response. The two news stories and their preconstructed responses are displayed in Table 1 and Table 2.

For Story 1, each of the responses index the conceptual metaphor **time is money**, which is well-entrenched in Western, English discourse and conceptualizes time as a resource which can be saved or spent (Lakoff & Johnson, 2008; Mueller, 2016). Because this entrenched metaphor is present in all four of the preconstructed responses, any differences in perceptions of creativity among these answers will likely be related to other features of the preconstructed responses. The first statement representing the **baseline** contains only this lexicalized metaphor (**waste of time**). In this manner the baseline response directly and unambiguously expresses the speaker’s concern that the lawsuit is not worthy of judicial review. In the second statement, the meaning of the baseline response is converted into a sarcastic response while still indexing the same conventionalized conceptual metaphor (**spending time**). The sarcastic statement uses negative irony to exaggerate a feigned enthusiasm for the presence of the lawsuit, which appears frivolous in light of the unrealistically high award amount sought by the plaintiff and thus the sarcasm contains an underlying message of disapproval for the woman’s actions. The sarcasm is marked through a combination of an exaggerated approval intensifier (i.e., *so glad*) and a surface meaning which is approximately opposite of the baseline meaning.

**Table 1**  Story 1 and constructed responses with conditions

<table>
<thead>
<tr>
<th><strong>Woman Sues Starbucks Over Amount Of Ice In Drinks</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A Chicago woman has filed a $5 million lawsuit against the Starbucks Corporation claiming false advertising of 24-ounce iced drinks that allegedly only contain 14 ounces of liquid. What do you think?</em></td>
</tr>
<tr>
<td>1. <strong>Baseline</strong></td>
</tr>
<tr>
<td>2. <strong>Sarcasm</strong></td>
</tr>
<tr>
<td>3. <strong>Metaphor</strong></td>
</tr>
<tr>
<td>4. <strong>Sarcasm and Metaphor</strong></td>
</tr>
</tbody>
</table>
In the third statement (the metaphor condition), the woman in the news story is metaphorically compared to a parasite. The negative connotation associated with parasitic behavior is supplanted onto the woman and her lawsuit, suggesting her attempts to sue for such a large amount of money are a drain on the civic resources associated with the justice system, serving to convey additional layers of negative evaluation. Finally, the fourth statement (the sarcasm and metaphor condition) combines the responses from the second and third statements so that sarcasm and metaphor are used in tandem to express disapproval.

In the second news story, the baseline response contains no figurative language and directly expresses the opinion that Inky the Octopus is better in the sea than in the aquarium. In the second statement (the sarcasm condition), the same opinion is provided sarcastically using positive irony, with the speaker feigning favor for the cramped conditions over the natural home of the octopus (i.e., the sea) but with an underlying message of approval for Inky’s escape. In the third statement, (the metaphor condition), the aquarium tank is metaphorically compared to a prison cell, again revealing additional layers of negative evaluation beyond the baseline and sarcasm conditions. In a similar fashion to the first story, the fourth statement combines the sarcasm and metaphor conditions.

### 4.2 Procedure

The two news stories and four preconstructed news responses were input into the online survey software Qualtrics. The presentation of stories and
preconstructed responses were counterbalanced among 16 different possible combinations, with the order of the news stories randomly presented within each combination. The experiment was then posted to Amazon Mechanical Turk (AMT). To be eligible for this study, participants had to reside in the United States and possess a minimum AMT job approval rating of 99% for at least 10,000 completed jobs. Participants who met this requirement first read the informed consent, which described the nature and purpose of the research, ensured them that their data would remain confidential, and stated that they would receive $1.00 USD for their participation. Participants who agreed then completed a short demographic survey asking them for their age (in years), sex (female or male), first language, level of education (did not complete high school, completed high school, completed two-year degree, completed four-year degree, completed master's or equivalent professional degree, or completed doctoral degree), and socioeconomic status (participants were shown a picture of a ladder representing wealth in society and asked to choose where they were on that ladder, with options including the bottom, near the bottom, in the middle, near the top, and at the top). These demographic questions were asked in order to ensure a diverse range of participants (age and sex) and control for measures which may influence their language exposure and use (first language, education, socioeconomic status).

After completing the demographic survey, participants read the task instructions. These instructions explained that the goal of the task was to develop creative news reactions for two real news stories by pretending to be a 24-hour news channel host. Participants were told that the fictional news channel's ratings depended on their ability to hold a fictional audience's attention, and therefore their job relied on their ability to be creative. Accordingly, they were asked to be as creative as possible and told that their reactions should reflect something they would actually say during live television, approximately equivalent to one written sentence in length. A definition of creativity as a novel yet effective solution to a problem was provided to the participants.

The instructions further explained that participants should type exactly what they would say, as they would say it, without including any additional detail about their thoughts or reactions to the story. After reading these instructions, participants were shown the first news story (randomly chosen), which presented the headline and brief summary of the story above a blank text box to enter their answer. After they entered their response, participants then saw their answer displayed above one of the randomly selected preconstructed answers (i.e., the baseline, sarcasm, metaphor, or sarcasm and metaphor conditions). The instructions indicated participants were to choose whether their answer was more, less, or equally as creative when compared...
to the preconstructed answer. After making this decision, participants then repeated the task for the second news story. There were no time constraints placed on participants to complete their answers. This procedure is summarized in Table 3.

All of the participant data was manually checked and participants whose data demonstrated a lack of attention to the task (i.e., provided nonsense or blank answers) were replaced (11 participants were removed in this manner). As such, the final distribution of participants was slightly unbalanced for each version of the experiment, with a mean of 26 participants per survey version ($SD = 1.65$, Min = 24, Max = 30) and a total of 420 participants remaining (from a grand total of 431).

### Participants
Data from a total of 420 participants recruited from Amazon Mechanical Turk were used in the current study. The participants represented an approximate 50/50 split among males and females (females = 215, males = 205). The average age of participants (rounded up) was 38 ($SD = 11.05$, Min = 20, Max = 71). Seven of the participants indicated their first language was not English. All but one of the participants indicated they completed high school, with 228 (54.3%) of the participants completing a four-year college degree or higher. When asked to self-rate their socioeconomic status, 169 (40%) chose on or near the bottom, 226 (53.8%) chose in the middle, and 25 (6%) participants indicated they were near or at the top of the socioeconomic ladder.

### Figurative Language and Play in Participant Responses
Participant response data was analyzed for potential linguistic creativity in the form of figurative language or play. Any answer which required additional...
inferencing beyond the literal meaning of the words (Gerrig & Gibbs, 1988) and/or included overt play with the meaning or form of the answer was coded as being potentially creative. Perhaps unsurprisingly, participants employed a range of linguistic strategies which could be considered potentially creative. Participants made conceptual comparisons, exploited the polysemous meanings of words, raised ironic rhetorical questions, stung with sarcasm, rhymed and blended to coin new words, and more. From these varied strategies three general categories of potential creativity emerged: making comparisons, playing with form and/or meaning, and being ironic. The frequency of these strategies along with explanations and examples are presented in Table 4. As can be seen, the Starbucks story attracted more verbal irony whereas the Inky story attracted more comparisons. Wordplay was present in both stories, albeit to a slightly greater extent in the Starbucks story.

Participant answers coded as Making Comparisons were those which invited the reader to make a comparison between the news story and some other referent, fictional or real. This was done via conceptual comparisons, similes, or implied comparisons and references which highlighted salient aspects of the news stories. A common strategy for the Inky the Octopus story was to compare Inky to famous escape artists (“Now that is one Octopus that can give Houdini a run for his money.”) or prisoners (“Inky is a fugitive from justice!”), as well as to make reference to movies such as Finding Dory which feature an octopus who also escapes from a museum (“I think that Octopus is on its way to find its friend Dory.”). Comparisons made for the Starbucks Ice story (“Is your coffee shop robbing you of your liquid gold?”) were less common.

Answers coded as Verbal Irony were those which contained sarcasm, ironic rhetorical questions, hyperbole, or ironic analogy. This strategy was relatively

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Frequency of different strategies signalling potential creativity in participant responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>Starbucks</td>
</tr>
<tr>
<td>Making Comparisons</td>
<td>13</td>
</tr>
<tr>
<td>Verbal Irony</td>
<td>52</td>
</tr>
<tr>
<td>Wordplay</td>
<td>71</td>
</tr>
<tr>
<td>Total Answers with Potential Creativity</td>
<td>134 (32%)</td>
</tr>
</tbody>
</table>

Note: Individual strategy values include three answers for Inky and one answer for Starbucks which contained both wordplay and comparison, as well as one answer for Starbucks which contained both irony and comparison. The final total row only counts each instance once.
rare for answers made in response to the Inky the Octopus story (“So nobody say[sic] a LARGE Octopus with multiple legs slither its way across the floor?”). Verbal irony was more common for the Starbucks Ice story, with participants sarcastically mocking the woman’s intelligence (e.g., “Does she know what ice is made of?”), exaggerating the stereotype of Starbucks coffee being expensive (“If she wins the lawsuit, she can buy 7 frappuccinos with her winnings.”), and finding analogies between the woman’s claims and similar situations (“Next up, a man sues every potato chip company ever for false advertising.”).

Wordplay was employed in response to both stories, with strategies including highlighting double meaning through polysemy, blending words together, making rhymes, or otherwise playing with the form and/or meaning of words. Several participants played with the polysemous meanings associated with hot and cold to make puns in response to the Starbucks Ice story, indexing conceptual metaphors related to temperature and emotion (e.g., “You could say she’s giving Starbucks the cold shoulder;”, “That Chicago woman is ice cold!”). Participants also played with meanings associated with coffee (“A woman has decided to get bold and sue Starbucks over ice.”) and the name of Starbucks itself (“This woman wants more drink for her buck at Starbucks.”). For Inky, participants focused on meanings related to arms and hands (“Sounds like that octopus was ‘well armed’”), physical features of Inky and other octopi (“Well that sure was a slippery escape!”), making rhymes (“Inky is slinky.”), and drawing attention to form overlap (“I have an INKling that the octopus is happier now.”)

It should be noted that these answers were not considered to be inherently creative but rather represented explicit attempts by the participants to do something with their answers that contained the potential to be considered linguistically creative. Moreover, the categories explained above reflected the primary strategy in the answer, but many answers contained elements of more than one type of figurative language or play (e.g., using polysemy to create a pun also indexed conceptual metaphors related to temperature and emotion). Therefore, in order to include this data in the statistical analysis, participant answers were coded as having or not having the potential for linguistic creativity (yes/no). In all, a total of 301 (35.8%) responses were found to contain the potential for linguistic creativity. The distribution was skewed slightly towards one story, with 167 (40%) of the answers for the Inky the Octopus story and 134 (32%) of the answers for the Starbucks Ice story containing the potential for linguistic creativity. The entire set of participant answers with codes is provided as Supplemental Material.2

2 https://osf.io/mvshr/.
4.5 Statistical Analysis
The dependent variable for this analysis is whether participants self-rated their response to each news story to be less creative, equally creative, or more creative when compared to the randomly selected, preconstructed response supplied to them. Because the dependent variable was categorical with three levels, an ordinal regression procedure was used. Specifically, a cumulative link mixed model (clmm) was fit, which is an ordinal regression allowing for individual intercepts to be fit for each participant and item (here, items refer to the eight different preconstructed responses). This approach is similar to a logistic regression predicting a binary outcome using mixed effect modelling, but instead predicts the probability of three outcomes, rather than two. As such, the interpretation of coefficients and odds ratios for ordinal regression is slightly different, in that the chances of predicting a higher level category is compared against the probability of all lower levels combined (Field, 2013; Levshina, 2015). In this case, positive coefficients and odds ratios greater than 1 for any variable indicates that variable is associated with a higher probability of choosing the more creative response when compared to the combined probability of choosing the equally creative or less creative response. Conversely, negative coefficients and odds ratios lower than 1 would reflect the opposite interpretation, representing a lower probability to choose the more creative response option when compared to the other two options.

The model was built in R version 4.0.0 (R Core Team, 2020) using the clmm function from the ordinal package, version 2019.12.10 (Christensen, 2019). The model contained participant self-ratings (less, equal, or more creative) as the dependent variable. The primary predictors of interest included pre-constructed response condition, which was a factor with four levels (baseline, sarcasm, metaphor, or metaphor and sarcasm), task, which was a factor with two levels (Inky the Octopus or Starbucks Ice), and whether the response had the potential for linguistic creativity, a factor with two levels (yes or no). The remaining predictor variables were included to account for demographic differences among the participants and were calculated from the answers provided to the demographic survey. These variables were participant age (in years), whether they were a native English speaker (yes or no), sex (female or male), socioeconomic status (bottom, middle, or top), and whether the participants had a four-year college degree (yes or no). Finally, random intercepts were fit for both participants and items.

The model fitting procedure began with testing the significance of random effects, then main effects, and then interactions between all main effects and
the preconstructed response condition. In all cases, log likelihood model comparisons between nested models were conducted using the ANOVA function in R. Only significant terms were retained in the final model. Post-hoc comparisons for categorical variables including more than two levels were conducted using the emmeans package, version 1.4.3.01 (Lenth, 2020). Odds ratios for terms in the models were calculated through exponentiation of the coefficients, providing a measure of effect size for each predictor variable (Levshina, 2015). Additionally, 95% confidence intervals for the odds ratios were computed using the confint function from the base stats R package.

5 Results

5.1 Participant Self-Ratings of Creativity
The distribution of participant self-ratings of creativity in the four different response conditions is depicted in Figure 1 and Figure 2. Figure 1 displays the overall distribution of self-ratings in the four different response conditions, whereas Figure 2 shows the distribution in the response conditions for each of the two news stories. As can be seen, the more creative category was selected more frequently when compared to the equally creative and less creative options, while the equally creative category was selected more frequently when compared to the less creative category. In other words, there is a visible tendency for participants to rate their responses as more or equally creative.

However, the distributions of these frequencies visibly differ among the different response conditions as well as between the news stories. For instance, the overall percentages for the more creative option decrease while percentages for the less creative option increase for the sarcasm, metaphor, and sarcasm and metaphor conditions when compared to the baseline condition (Figure 1). Moreover, the frequency of self-ratings in the four response conditions is different when comparing the two news stories (Figure 2). For example, no participants indicated their responses were less creative to the baseline condition for the Starbucks Ice story, whereas 14 participants chose this option for the baseline condition response for the Inky the Octopus story. In other words, Figure 1 and Figure 2 visibly suggest an interaction between response condition and the two different news stories.

5.2 Potential Linguistic Creativity and Self-Ratings
The distribution of participant self-ratings is also plotted in light of whether their answer was coded for the presence or absence of potential linguistic creativity in Figure 3. When comparing the differences among the three ratings
Figure 1: Raw counts of participant self-ratings of creativity separated by preconstructed response condition.

Figure 2: Raw counts of participant self-ratings of creativity separated by preconstructed response condition and news story.
choices for answers that did and did not contain potential linguistic creativity, Figure 3 visually suggests a greater tendency for participants to self-rate their answers as *more creative* when their answers also contained the potential for linguistic creativity. This is demonstrated by the higher absolute percentages for the *more creative* option and lower absolute percentages for the *less creative* option for responses that were coded as *yes* when compared to those coded as *no*. For the most part this effect is visually consistent, with some evidence of a difference in distribution for certain conditions, suggesting a potential interaction between the presence of potential linguistic creativity and response condition.

### 5.3 Ordinal Regression

Results from comparisons to an ordinal regression model with all variables fit as main effects found that random intercepts fit on items (i.e., the eight different preconstructed responses) contributed close to zero variance and did not significantly improve model fit ($\chi^2(1) < .001$, $p = .984$) and were thus excluded from subsequent models. The random intercepts fit for subjects did explain a significant amount of variance ($\chi^2(1) = 17.33$, $p < .001$) and were retained. Further results from comparisons to a model with all variables fit as main effects and random intercepts fit for subjects found that the main effects of response condition ($\chi^2(3) = 20.184$, $p = .001$), news story ($\chi^2(1) = 9.651$, $p = .002$), potential linguistic creativity ($\chi^2(1) = 63.457$, $p < .001$), and participant age ($\chi^2(1) = 6.546$, $p = .011$) were all significant improvements compared to models without those effects. On the other hand, socioeconomic status ($\chi^2(2)$)
native English-speaking status ($\chi^2(1) = 0.314, p = 0.575$), holding a four-year college degree ($\chi^2(1) = 1.093, p = .296$), and participant sex ($\chi^2(1) = 1.364, p = .243$) did not significantly improve the model fit and were thus removed from the model as main effects.

Finally, results from comparisons to a model with the significant random and main effects listed above found that an interaction between response condition and news story contributed to a significantly better model fit ($\chi^2(3) = 11.083, p = .011$). Interactions between response condition and potential linguistic creativity ($\chi^2(3) = 2.461, p = .482$), age ($\chi^2(3) = 0.790, p = .852$), holding a four-year college degree ($\chi^2(4) = 1.384, p = .847$), socioeconomic status ($\chi^2(8) = 5.038, p = .754$), and sex ($\chi^2(4) = 2.391, p = .664$) did not result in a significant model fit and were thus not retained, and a model with an interaction between English-speaking status and response condition failed to converge. Accordingly, the final model included a random effect of subject, a significant main effect for potential linguistic creativity in the participant responses, a significant main effect for participant age, and a significant interaction between response condition and news story (of which both were also significant main effects).

5.3.1 Main Effect of News Story and Response Condition

The significant main effect of news story predicted a 1.64 times higher likelihood for participants to choose the *more creative* self-rating option in response to the Starbucks Ice news story when compared to the Inky the Octopus news story (OR = 1.640, 95%CI [1.252, 2.147], $p = .003$). In terms of response condition, post-hoc analyses indicated that compared to the baseline condition, participants were significantly less likely to choose the *more creative* self-rating option for all three of the other preconstructed answer types: sarcasm (OR = 1.749, 95%CI [1.141, 2.680], $p = .031$), metaphor (OR = 2.625, 95%CI [1.720, 4.005], $p < .001$), or sarcasm and metaphor (OR = 2.696, 95%CI [1.750, 4.154], $p < .001$). There were no significant differences in self-rating likelihood among the sarcasm, metaphor, and sarcasm and metaphor response conditions.

The main effects of news story and response condition are displayed in Figures 4 and 5, respectively. In these figures, the probability of choosing each of the three self-rating options is measured along the y-axis. The different response conditions and news stories are plotted along the x-axis based on their logit odds extracted from the regression model. The curved horizontal lines represent the variation in probability for each of the three self-rating options, whereas the vertical dotted lines represent the values for either the different news stories (Figure 4) or preconstructed response types (Figure 5). As can be seen in Figure 4, the probability associated with choosing the *more*
A more creative rating was significantly more likely for the Starbucks Ice story when compared to the Inky the Octopus story (OR = 1.640, 95% CI [1.252, 2.147], p = .003).

A more creative rating was significantly more likely for the baseline condition when compared to the other response conditions: sarcasm (OR = 1.749, 95% CI [1.141, 2.680], p = .031), metaphor (OR = 2.625, 95% CI [1.729, 4.005], p < .001), sarcasm and metaphor (OR = 2.696, 95% CI [1.750, 4.154], p < .001). No other contrasts were significant.

The probability for the more creative option for the Starbucks Ice story is approximately 60%, whereas the probability for that same option is around 50% for the Inky the Octopus news story. The probabilities associated with the equally and less creative option are higher for the Inky the Octopus story and lower for the Starbucks Ice story.
Figure 5 displays the response conditions along the same probability curves. The baseline condition is furthest right on the x-axis and is associated with the highest probability of choosing the *more creative* self-rating option, around 50%. All three of the other response conditions are further left on the x-axis and are associated with relatively lower probabilities of *more creative* and relatively high probabilities of *equally* or *less creative* self-ratings.

5.3.2 Interaction between News Story and Response Condition

The final ordinal regression model included a significant interaction between news story and response condition, suggesting the differences displayed in Figure 4 and Figure 5 depend on which news story the response conditions were associated with. Follow up post-hoc tests indicated this was indeed the case. For the Inky the Octopus news story, only the difference between the baseline and the combined sarcasm and metaphor response condition differed significantly, in that the baseline condition was associated with a 2.4 times greater likelihood of attracting a *more creative* rating when compared to the sarcasm and metaphor condition (OR = 2.400, 95%CI [1.324, 4.351], \( p = .015 \)). Conversely, for the Starbucks Ice story, the baseline condition had a significantly higher likelihood of *more creative* ratings when compared to metaphor (OR = 5.324, 95%CI [2.853, 9.935], \( p < .001 \)) and combined sarcasm and metaphor (OR = 3.185, 95%CI [1.698, 5.973], \( p = .002 \)) conditions. Moreover, the sarcasm condition in the Starbucks Ice news story was associated with a 3.2 times greater likelihood of a *more creative* rating when compared to the metaphor condition in the Starbucks Ice news story (OR = 3.158, 95%CI [1.735, 5.746], \( p = .002 \)).

In a similar fashion to Figures 4 and 5, the interaction between response condition and news story is plotted against the probability curves of the three self-rating conditions in Figure 6 and Figure 7. As can be seen, the four response conditions are grouped relatively closer for the Inky the Octopus story (Figure 6), whereas greater variation is seen for the Starbucks Ice story (Figure 7). The relative distance between the conditions reflects the significant contrasts described above. Overall, Figure 7 suggests the condition associated with the lowest probability of *more creative* ratings was the metaphor condition in the Starbucks Ice news story, around 14%. The baseline condition for the Starbucks Ice story was associated with the highest probability of *more creative* ratings, somewhere above 60%. 
5.3.3 Potential Linguistic Creativity in Participant Responses

The significant main effect of potential linguistic creativity in participant responses indicated participants were ~4.6 times more likely to choose the more creative option when compared to the equally and less creative options (OR = 4.644, 95%CI [3.248, 6.640], $p < .001$) if their answer was coded as yes for potential linguistic creativity. This effect is visually plotted in Figure 8.

5.3.4 Age

The significant main effect of age in the final regression model indicated a positive association between age and the predicted logit odds of choosing the more creative option when compared to the equally and less creative options. Specifically, the calculated odds ratio for this term predicted a 1.023 times greater likelihood of making this selection for each yearly increase in age (OR = 1.023, 95%CI [1.007, 1.038], $p = .016$). Recalling that the minimum and maximum age range in this data was 20 and 71 years old, the practical interpretation of this effect is that each yearly increase in age from 20 years to 71 years old adds an additional 1.2% likelihood of choosing the more creative option. Similar to the previous variables, the effect of age is visually plotted against the
Figure 7: Interaction effect of story and response condition on participant rating behavior for Starbucks Ice story. A *more creative* rating was significantly more likely for the baseline condition when compared to the metaphor (OR = 5.324, 95%CI [2.853, 9.935], $p < .001$) and combined sarcasm and metaphor (OR = 3.185, 95%CI [1.698, 5.973], $p = .002$) conditions. A *more creative* rating was also significantly more likely for the sarcasm condition when compared to the metaphor condition (OR = 3.158, 95%CI [1.735, 5.746], $p = .002$). No other contrasts were significant.

Figure 8: Main effect of potential for linguistic creativity in participant responses on participant rating behavior. A *more creative* rating was significantly more likely if participants included figurative language or play in their response (OR = 4.644, 95%CI [3.248, 6.640], $p < .001$).
probability curves of the three self-rating response options. Figure 9 displays where the minimum, maximum, and mean ages fall in this range, as well as one standard deviation from the mean in either direction.

6 Discussion

The purpose of this study was to investigate layperson perceptions of creativity and their potential relation to figurative language and play. To do so, human self-ratings of creativity were gathered from a creative production task. In this task, participants pretended to be news anchors for a 24-hour television news channel and wrote creative, attention-grabbing reactions for two different news stories depicting real yet difficult-to-believe events. For each of the two news stories, participants were shown their response alongside a randomly chosen, preconstructed response which contained varying amounts of sarcasm and metaphor. Participants then selected whether they felt their response was less creative, equally creative, or more creative than the preconstructed response. In addition, the participant responses were coded for potential linguistic creativity operationalized as figurative language or other explicit attempts at language play. The probabilities for self-rating selections were statistically modelled in order to evaluate whether the different types of preconstructed responses, the different news stories, potential linguistic creativity in participant answers, and other demographic features of the participants influenced their self-ratings of creativity.
6.1 *Influence of Preconstructed Responses on Self-Ratings of Creativity*

The first research question asked whether self-ratings of creativity were influenced by the presence of figurative language in the different preconstructed responses provided to the participants for comparison. The results obtained from the main effects ordinal regression suggest that preconstructed answers containing metaphor or sarcasm were associated with lower probabilities for participants to choose the *more creative* rating (when compared to the baseline condition). Although participants provided ratings for their own answers, these ratings were also comparisons and thus indirectly indexed perceptions of creativity towards the preconstructed answers. In other words, this would suggest that the statements containing figurative language were perceived to be more creative than those that did not contain figurative language because the baseline answer was significantly more likely to attract a *more creative* comparison from the participants when compared to the answers containing metaphor and/or sarcasm. However, as seen in Figure 6 and Figure 7, this main effect interacted with both story and response condition, suggesting further influences beyond the presence or absence of figurative language in the response condition.

6.1.1 Interaction with Story Condition

Although results from the main effects regression model reported that all non-baseline response conditions were associated with a significantly lower probability of *more creative* self-ratings from the participants, a closer look at this effect indicated the self-ratings were also influenced by the different stories. Crucially, the significant difference between the *sarcasm only* and the *baseline* response conditions was no longer observed when separating the statistical results based on the two stories. Moreover, the significant difference between the *metaphor only* and *baseline* response condition remained only for the Starbucks Ice story, while only the contrast between the combined *sarcasm and metaphor* and *baseline* conditions remained significant for the Inky the Octopus story. These results suggest the figurative language used in preconstructed responses for the Starbucks Ice story had a different influence on self-ratings of creativity when compared to preconstructed responses for the Inky the Octopus story.

6.1.2 Starbucks and the Parasite Metaphor

The primary difference between the stories is that the *metaphor only* condition for the Starbucks story ("This woman is a parasite wasting the time of our judicial system") was associated with a lower probability of receiving a *more creative* self-rating when compared to the *baseline* condition, whereas the same was not true for the *metaphor only* condition in the Inky story. This difference
was also associated with the strongest statistical effect size, with an odds ratio of 5.32. Moreover, the combined *sarcasm and metaphor* condition from the Starbucks story (“*I’m so glad our legal system is spending time on parasites like this*”) included the same metaphorical comparison and had the second strongest effect size, with an odds ratio of 3.18. These effects cohere to suggest that this specific metaphor, which framed the woman in the Starbucks story as a parasite, was far less likely to garner *more creative* self-ratings from the participants when compared to the *baseline* condition (and thus was perceived to be more creative when compared to other preconstructed responses).

The parasite metaphor has historically been used to frame groups of people (such as immigrants) as slowly destroying other entities (such as countries) from within (Musolff, 2012). By shifting this metaphorical description to the overly litigious woman in the Starbucks Ice story, the answers containing the parasite metaphor framed the woman as harming the United States from within, indexing a general cultural disapproval for excessive litigation in the United States (Rhode, 2004). Moreover, because the parasite metaphor has typically been applied to groups of people, the use of the metaphor towards a single person may have appeared more extreme or more noticeable for the participants, which in turn may have been perceived as more novel (and thus, more creative). In contrast, the Inky the Octopus story was more positive, constructing a feel-good atmosphere of an intelligent animal beating the odds to escape, and the preconstructed responses all worked to celebrate the actions of Inky. The metaphor in the Inky story, equating the aquarium to a prison cell, may have been unsurprising because the story framed the event as an escape, and thus any comparisons to a prison cell may not have seemed overly novel or original and thus not as creative.

### 6.1.3 Inky and Sarcasm

Unlike the Starbucks story, results obtained for the Inky story suggest that self-ratings of creativity were only significantly different from the baseline condition when participants were faced with a combination of both sarcasm and metaphor (“*I’m sure the sea will be much worse compared to his prison cell.*”). In terms of the size of this effect, the odds ratio from the regression model was 2.40 (i.e., 2.4 times greater likelihood to choose *more creative* when making a comparison against the baseline condition rather than the combined *sarcasm and metaphor* condition). While not as strong as some of the other effects obtained for the Starbucks Ice story, this effect was still significant and suggests that there was something different about this particular combination of sarcasm and metaphor. One potential explanation for this effect is that the
The second research question asked whether or not participants would include figurative language in their answers when explicitly asked to be creative, and whether any inclusion of figurative language would influence self-ratings of creativity. The results obtained here suggest participants did indeed include figurative language in their answers, but also that participant answers included strategies which transcended the boundaries of just being metaphorical or just being sarcastic. In all, figurative language or some other form of language play occurred in approximately one third of the provided answers. Although this percentage might seem relatively low, the overall likelihood for these answers to be associated with a more creative rating was statistically strong: participants were 4.6 times more likely to rate their answer as more creative if they included figurative language or play in their response. Considered in light of the odds ratios for other effects, this was overall a relatively strong effect. Moreover, this effect did not significantly interact with story (Inky vs. Starbucks) or the four preconstructed response conditions, suggesting this effect was stable regardless of whether participants were making comparisons against answers that did or did not contain metaphor or sarcasm. As such, these results provide further and direct evidence that perceptions of creativity are associated not just with figurative language but also with more general forms of language play.

Although not incorporated into the statistical analysis, Table 4 suggests differences in the types of figurative language and play based on the news story. In general, the Inky story attracted more comparisons whereas the Starbucks story attracted more irony, with both stories eliciting a somewhat balanced number of wordplay responses. These differences are likely attributed to differences in the tone and content of the news stories. The Starbucks Ice story contained a more fertile context for sarcasm and verbal irony by providing

---

3 Many thanks to the anonymous reviewer who brought this potential explanation to my attention.
multiple targets for which participants could express sarcasm towards, such as the woman (“Woman forgets that Ice melts.”), the inflated price of coffee from Starbucks (“Wow, 5 million dollars could get her another 100 or so coffees from Starbucks!”), and the nature of frivolous lawsuits (“Now if I could just sue McDonald’s for the amount of hair in my Big Macs.”). Conversely, the setting of the Inky story facilitated comparisons to other well-known escapes, including movies (“Like a scene from Finding Dori, Inky the Octopus squeezed out of tank and into freedom.”) and real prisons (“Octopus would put Alcatraz to shame.”). Finally, it is likely that further differences exist among the participant answers in that some are more figurative, more playful, or more creative than others if measured using rubrics or metrics associated with theoretical definitions of figurative language, language play, or creativity. Regardless of these differences, the current results provide evidence demonstrating that the explicit attempt by participants to be figurative or to play with language was enough to significantly influence their self-perceptions of creativity.

6.3 Participant Age
A relatively unexpected main effect was also obtained for age, in that increases in age were associated with increases in participants selecting their own responses to be more creative. This effect did not interact with any response or task condition, meaning that this effect applied regardless of the different levels of figurative language in the preconstructed responses. One potential explanation for this effect is that increased age brings increased exposure to language, and thus older participants may have further found the metaphors and sarcastic utterances used in the preconstructed responses to be less novel and therefore less creative. Older participants may also have been more confident in their ability to perform the task and thus tended towards the more creative option. These explanations are both partially supported by one study which demonstrated that aging artists had higher self-perceptions of creativity, with the explanation that increased experiences associated with a longer life provide more inspirations for creative performance (Lindauer et al., 1997). It may be the case that a similar phenomenon is at play in the current data, with older participants naturally assuming a heightened sense of creativity. Further research into the connection between age and perceptions of creativity associated with language use are thus warranted.

7 Conclusion
The goal of this study was to further explore connections between creativity and figurative language use. It is important to note that the nature of the task
used here is just one way of measuring perceptions of creativity. A simpler yet more straightforward approach to measuring such perceptions of creativity in language could be to ask participants to rate a series of statements for their creativity, which would not involve a comparison to one’s own creative ability and may avoid any potential face-threatening thoughts associated with self-evaluation. However, by asking participants to craft their own responses and then compare them to preconstructed responses, the current data allows for an increased understanding into the subconscious considerations people make when asked to evaluate the creativity of language.

Overall, the results reported here do suggest that metaphor has a greater potential to be perceived as creative when compared to sarcasm. However, in the same manner as results from research on elicited figurative language production, these results appear to depend on the specific type of metaphorical comparison being made. In this case, the woman as parasite metaphor influenced self-ratings more strongly than the aquarium as prison metaphor. Sarcasm, on the other hand, did not appear to be strongly associated with increased perceptions of creativity. An additional and notable finding from this research is that participants appeared to hold their answers in high regard, with a general tendency to rate their own answers as more creative than any of the preconstructed answers. This effect was significantly amplified if participants employed figurative language or language play in their own response, suggesting this to be a salient strategy associated with writing creative responses. Because of the strong effect found for the inclusion of figurative language or play in the participants’ responses, future research in this area may want to consider play as another response condition, as well as expand this task to a greater number of prompts and contexts. In all, the layperson perspective represented in the current data serves to bolster academic arguments exploring the role between creativity, figurative language, and play.

Acknowledgements

Many thanks to Scott Crossley for kindly funding the data collection.

References


