Exploring Perceptions of Novelty and Mirth in Elicited Figurative Language Production

Stephen Skalicky

Routledge
Taylor & Francis Group

Check for updates

Exploring Perceptions of Novelty and Mirth in Elicited Figurative Language Production

Stephen Skalicky

Victoria University of Wellington

ABSTRACT

Most research of figurative language production examines naturalistic discourse. However, laboratory studies of elicited figurative language production are useful because they provide insight into whether specific individual differences are associated with differences in figurative language production ability. In this sense, elicited figurative language production studies mirror the approach many studies of figurative language processing and comprehension take. Accordingly, this study is an investigation into perceptions of novelty and mirth associated with figurative language production. Data in this study come from elicited figurative language production tasks for metaphorical comparisons and sarcastic replies, which were rated based on perceptions of novelty and mirth as well as conceptual distance (metaphors) and incongruity (sarcastic replies). Two individual differences, Need for Cognition and Abstract Thinking, along with other differences such as prompt type and production time were used to predict the novelty and mirth ratings. The results demonstrate a significant effect of Need for Cognition on metaphors but not for sarcastic replies, whereas no significant effects were observed for Abstract Thinking. Moreover, longer production times were associated with higher ratings for both metaphors and sarcastic replies. Finally, the results highlighted the manner in which prompt characteristics influenced participants' responses, especially so for sarcastic replies.

Introduction

When one surveys the existing empirical research into figurative language use, several trends tend to emerge. For instance, much of the scientific knowledge of figurative language comes from psycholinguistic experiments examining the online processing and comprehension of metaphor, verbal irony, and other types of figurative language. These studies usually test different theoretical predictions regarding the role of literal and figurative meaning during figurative language processing and comprehension (e.g., Filik, Leuthold, Wallington, & Page, 2014; Gibbs, 1986; Gibbs & Colston, 2012; Giora, 2003). Additionally, studies from this perspective have identified associations between specific cognitive individual differences and figurative language processing. Some of these differences include working memory capacity, self-reported sarcasm use, or the desire to engage in cognitively difficult tasks (Ivanko, Pexman, & Olineck, 2004; Kaakinen, Olkoniemi, Kinnari, & Hyönä, 2014; Olkoniemi, Johander, & Kaakinen, 2018; Olkoniemi, Ranta, & Kaakinen, 2016; Skalicky, 2019).

Another rich source of knowledge has been generated from research investigating figurative language production. The literature in this area suggests that figurative language production has been extensively investigated in a wide variety of communicative contexts, including scientific

CONTACT Stephen Skalicky ✉ scskalicky@gmail.com 🖃 School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand

discourse (Deignan, Littlemore, & Semino, 2013), university lectures (Corts & Meyers, 2002; Low, Littlemore, & Koester, 2007), language learner corpora (Nacey, 2013), e-mails (Whalen, Pexman, & Gill, 2009), casual conversation (Gibbs, 2000; Hussey & Katz, 2006), and more (Gibbs & Colston, 2012). Perhaps unsurprisingly, this research suggests that speakers use figurative language for a variety of discourse functions and goals (Colston, 2015; Gerrig & Gibbs, 1988; Gibbs, 1994; Roberts & Kreuz, 1994), and also that personality, occupation, relationship status, and other speaker characteristics influence expectations and perceptions associated with figurative language production (Hussey & Katz, 2006; Katz & Pexman, 1997).

What is less common are studies that include prompts explicitly asking participants to produce examples of figurative language in controlled conditions. In a similar manner to some of the more recent figurative language processing studies mentioned above, elicited figurative language production studies have identified associations among figurative language production and specific individual differences, including intelligence, working memory capacity, abstract thinking, and creative ability (Beaty & Silvia, 2013; Chiappe & Chiappe, 2007; Huang, Gino, & Galinsky, 2015; Pierce & Chiappe, 2008; Silvia & Beaty, 2012). However, when compared to the number of studies testing figurative language processing, elicited figurative language production studies are relatively few. Perhaps this is because naturalistic, spontaneous figurative language production is more representative and authentic. Nonetheless, elicited production studies are necessary in order to continue investigating potential associations among figurative language production and individual differences because these participant-level measures cannot be gathered alongside some observational kinds of studies investigating spontaneous, naturalistic production as found in corpora and other existing datasets of figurative language.

Accordingly, the research reported in the current study is from an elicited production approach designed to further investigate associations among individual differences and figurative language production. Specifically, two types of figurative language were collected: metaphorical comparisons and sarcastic replies. The definition of metaphor used here is inclusive in that metaphorical comparisons and similes are treated as approximately similar constructs because they are both used to explain one thing in terms of another (Haught, 2013). The definition of sarcasm used here is a stereotypically negative form of verbal irony, wherein a speaker produces an utterance meant to express their (typically) negative or pessimistic opinion about someone or something. In order for an utterance to be sarcastic (and thus ironic), the surface message of the utterance must somehow clash or be incongruous with the situation or context the utterance refers to (Attardo, 2000; Colston, 2017).

Perceptions of originality, mirth, and conceptual distance between entities (metaphors) or incongruity between meaning and context (sarcastic replies) were then obtained and compared in light of two individual differences. The first, Abstract Thinking (AT), measures how people perceive and identify different events in terms of immediate or future implications of an action (Vallacher & Wegner, 1989). Specifically, if someone possesses higher levels of AT, they tend to identify larger, more abstract meanings of an action, which include motivations and future goals of the action. For example, the action of reading can be described either as *following lines of print* (concrete) or *gaining knowledge* (abstract). The second, Need for Cognition (NFC), measures an individual's desire to engage in tasks that require high levels of cognitive effort (Cacioppo, Petty, & Kao, 1984). NFC is measured by asking participants to answer how strongly they agree or disagree with statements such as *I like tasks that require little thought once I've learned them*. As will be discussed below, both AT and NFC have been previously associated with either metaphor processing or sarcasm use. However, they have not yet been thoroughly investigated in both metaphor and sarcasm production. The following section reviews prior elicited figurative language production research.

Elicited metaphor production

The results of four studies represent most of the knowledge in this area (Beaty & Silvia, 2013; Chiappe & Chiappe, 2007; Pierce & Chiappe, 2008; Silvia & Beaty, 2012). The goals of these studies were similar

in that they all sought to investigate whether different measures of executive (cognitive) control were related to metaphor production ability (and by extension, creative ability). The earlier two studies investigated working memory capacity (the ability to perform complex operations while holding relevant information stable in short-term memory), but also measured vocabulary knowledge (Chiappe & Chiappe, 2007; Pierce & Chiappe, 2008). In these studies, participants were presented with a topic and a description of that topic (e.g., *music can be something that heals and makes one feel better*). They were then asked to produce a metaphor or simile which possessed the same meaning as the provided description. Measures of metaphor production quality were then obtained through human ratings. In Chiappe and Chiappe (2007), metaphor quality was operationalized as aptness of the metaphorical vehicle chosen by participants, and results included positive correlational associations between metaphor quality and both working memory capacity and vocabulary knowledge (Chiappe & Chiappe, 2007). In Pierce and Chiappe (2008), separate ratings were conducted for quality, aptness, and conventionality. Results from this study suggested that both working memory capacity and conventionality ratings were positive predictors of metaphor quality ratings.

The other two studies examined different measures of intelligence to investigate links between intelligence, metaphor production, and the cognitive construct of creative ability (Beaty & Silvia, 2013; Silvia & Beaty, 2012). These studies both measured fluid intelligence (problem-solving ability), while Beaty and Silvia (2013) also measured crystallized intelligence (background knowledge) and broad retrieval (in this case, vocabulary knowledge). Additionally, these studies made a further distinction between conventional (clichéd or previously encountered) and novel (original) metaphors. Accordingly, Silvia and Beaty (2012) instructed participants to produce metaphors in response to two past experiences (i.e., the most boring class they had been in and the most disgusting food they had eaten) based on the presumption that biographical memory is more suitable for eliciting original metaphorical comparisons. Participants in Beaty and Silvia (2013) performed the same novel metaphor task and also produced metaphors in response to the same stimuli used in Chiappe and Chiappe (2007), which Beaty and Silvia (2013) argued were more likely to capture conventional metaphors. In both the 2012 and 2013 studies, measures of metaphor quality were obtained through human ratings of creativity, based on a holistic perception of metaphor novelty, remoteness of concepts, and cleverness. While results from these studies included associations among all three measures of intelligence and ratings of metaphor creativity, differences were found in that fluid intelligence and broad retrieval were more strongly associated with creativity ratings of the novel metaphors, whereas crystallized intelligence was more strongly associated with the conventional metaphors.

Several conclusions can be drawn from these results. The studies of metaphor production do suggest that executive control is associated with metaphor production ability, and that vocabulary knowledge is also important. The results further suggest participants may draw upon different cognitive resources when asked to produce novel versus conventional metaphors. Specifically, quality ratings of novel metaphors were associated with cognitive measures of problem-solving, whereas conventional metaphors were associated with preexisting knowledge. These findings make intuitive sense in that participants can rely more strongly on prior experience to evoke previously encountered metaphorical comparisons when completing the conventional metaphor prompts. When constructing original, novel metaphors, participants instead drew upon cognitive resources more associated with creative problem solving. Moreover, although not the primary question in any of the metaphor production studies, participants who spent longer times creating metaphors tended to produce metaphors with higher quality ratings.

Elicited verbal irony production

There are even fewer studies of elicited verbal irony or sarcasm production. For instance, while Ivanko et al. (2004) reported individual variation in sarcasm production, sarcasm production in that study was measured by asking participants to choose among a set of four pre-selected answers to prompts, with one of the answers reflecting a sarcastic intention. As such, participants did not produce their

own individual sarcastic responses. More recently, Huang et al. (2015) investigated potential associations among sarcasm use, abstract thinking, and creative ability. Their study investigated the perception, recall, and production of sarcastic remarks in a series of four separate experiments. For sarcasm production, some participants were shown black-and-white cartoons depicting frustrating situations and asked to write the first sarcastic reply that came to their minds (Experiments 1 and 4), while others were asked to recall an incident in which they had previously said something sarcastic (Experiments 2 and 3). Overall, Huang et al. (2015) argued that sarcasm use (including production) increased levels of abstract thinking for participants, which in turn led to increased performance on different tests of creative ability. Huang et al. (2015) concluded that both sarcasm and creative ability were linked by a similar construct: abstract thinking.

Figurative language and creativity

A central concept linking the research of both metaphor and sarcasm production reviewed above is the notion of creativity and how it is (or is not) associated with different cognitive individual differences. Creativity has long been associated with figurative language use (Gerrig & Gibbs, 1988). As a cognitive construct, creative ability is described as an individual difference in the same manner as intelligence or working memory capacity. Creative ability is commonly operationalized as one's aptitude to develop novel and effective solutions to problems (Runco & Jaeger, 2012). However, figurative language use is not restricted to individuals with enhanced creative ability, and some examples of figurative language use are formulaic and standardized in their use (Veale, 2012). Nonetheless, as the research reviewed above has suggested, individuals can differ in terms of how creative their figurative language is perceived to be by others. As such, perceptions of creativity in figurative language may provide some insight into figurative language production because these perceptions are associated with specific participant individual differences. Therefore, it is worthwhile to continue investigating cognitive differences that may be associated with creative processes, such as figurative language production. In the current study, two specific cognitive traits are studied: Abstract Thinking and Need for Cognition.

Abstract thinking

In a study reviewed above, Huang et al. (2015) explained potential links between sarcasm and creativity by citing theoretical descriptions of sarcasm, which defined sarcasm as an utterance which "signifies the opposite" (p. 163) and thus results in contradictions between meanings. Accordingly, Huang et al. (2015) postulated that the ability to develop and consider multiple meanings is inherently a creative ability, and therefore using sarcasm is a creative process that benefits from (and induces) higher levels of abstract thinking. While this view of sarcasm relies on simplistic definitions of verbal irony and sarcasm that may not accurately describe the nature of figurative meaning (Gibbs & Colston, 2007, 2012), the overall interpretation that sarcasm use is a creative process may nevertheless provide a beneficial theoretical lens through which to examine sarcasm and other types of figurative language.

Need for cognition

Need for Cognition (NFC) is one feature that has been explored during figurative language processing but has yet to be investigated in terms of figurative language production. NFC is described as a measure of one's desire to perform tasks which are cognitively complex or demanding (Cacioppo et al., 1984). The prevailing hypothesis is that if figurative language requires more cognitive effort to process and comprehend, then participants with high levels of NFC should be better able to process and comprehend figurative meaning. However, studies have shown that the predictions of this hypothesis are not always accurate. For example, no connection has been found between NFC and verbal irony processing (Kaakinen et al., 2014; Olkoniemi et al., 2016). These same studies did however identify a link between higher NFC and metaphor processing, but found that participants with higher

NFC took significantly longer to process metaphors, counter to the expectations that NFC should facilitate processing (Olkoniemi et al., 2016). Finally, a different study found that higher levels of NFC predicted faster reading times for satirical compared to non-satirical discourse (Skalicky, 2019). As such, the role of NFC during figurative language processing appears complex and is likely dependent in part on the type of figurative language under investigation.

Current study

The above review suggests there is a need to further test individual differences which have previously been associated with figurative language processing and comprehension within the context of figurative language production. If the same features associated with figurative language processing are also associated (or not associated) with figurative language production, this would provide stronger evidence of a link between figurative language use and specific cognitive individual differences, including creative ability. The studies reviewed above have also suggested that vocabulary knowledge and time spent producing figurative language are associated with perceptions of creativity or figurative language quality and thus also warrant further study. Finally, and more practically, there is a general need for more studies of elicited figurative language production as most of the scientific research of figurative language is focused on processing and comprehension, or is focused on figurative language produced spontaneously in naturalistic discourse. Accordingly, the primary purpose of the current study is to further investigate metaphor and sarcasm production as measured through human perceptions of creativity in light of two individual differences: Abstract Thinking and Need for Cognition.

Method

Participants

Sixty-one undergraduate and graduate students from a large public university in the United States were recruited to participate in the study. Participant ages ranged between 17 and 63 years old (m = 25.56, SD = 8.341), and 46 were female and 15 were male. Approximately half (n = 32) of these participants were international students enrolled in the university and did not speak English as their first language, whereas the remainder (n = 29) all indicated English to be their first language. The entry requirements for international students at this university required English proficiency scores typically attributed to a B2 level using the Common European Framework of Reference for Languages (CEFR). The CEFR description and core inventory of functions for B2 level English users include describing experiences, feelings, and emotions using advanced syntax and colloquial language (North, Ortega, & Sheehan, 2011).

Materials

Need for cognition
Participants' Need for Cognition was measured using a previously tested and validated survey (Cacioppo et al., 1984). The survey includes 18 items total and instructs participants to rate agreement or disagreement with different statements designed to capture an innate desire to perform complex cognitive tasks. Participants used a scale ranging from −4 (very strong disagreement) to 4 (very strong agreement) to indicate agreement with statements such as *I only think as hard as I have to*. Final scores are the result of the sum of all agreement answers, which can range from −72 (low NFC) to 72 (high NFC).

Abstract thinking
Participants' Abstract Thinking (AT) was measured using the Behavior Identification Form (BIF; Vallacher & Wegner, 1989). The BIF is a list of 25 actions (e.g., *Making a list*) accompanied with two

descriptions of that action, one representing a concrete interpretation (e.g., *writing things down*) and one representing an abstract interpretation (e.g., *getting organized*). For each action, participants select the description they feel is the best interpretation of that action. The total AT score is the sum of all the time participants choose the abstract interpretation of an event, for a total possible score of 25.

Metaphor production prompts

Two different metaphor production tasks were developed using stimuli from metaphor production studies reviewed above (Beaty & Silvia, 2013; Chiappe & Chiappe, 2007). The first type of prompt was a conventional metaphor task and contained 22 different items. Each item included a Topic and a Description. All of the Topics consisted of noun phrases (e.g., her family), and all of the Descriptions were descriptions or properties of those noun phrases (e.g., something that keeps her stable and prevents her from drifting into danger). The goal for each prompt was to generate a metaphorical description of the Topic reflective of the same meaning in the provided Description, but without repeating any of the words from the Description. This task allows participants to draw upon typical or previously encountered metaphorical comparisons they may already be familiar with, hence its description as a conventional metaphor task.

The second type of prompt was a novel metaphor task also based on prior studies. This task asked the participant to recall their experience with two scenarios: the most boring class they have been in, and the most disgusting item they have ever eaten or drank. For each scenario, participants were instructed to produce a metaphor that described their feelings during that scenario and were also provided with an example of how to start their metaphors (e.g., Being in that class was like _____). This prompt provided more freedom to make original and novel metaphorical comparisons based on each individual's lived experience.

Sarcasm production prompts

Twelve different drawn cartoons were adapted or created to serve as sarcasm production prompts. Three different types of cartoons were created with four cartoons per type. The first type consisted of black and white cartoons used by Huang et al. (2015). These four cartoons were selected from the Rosenzweig Picture Frustration Study, originally designed to assess patient responses to frustrating situations in order to diagnose aggression (Rosenzweig, 1945). Each of the black and white cartoons is a single-panel cartoon which depicts a frustrating situation involving more than one speaker (e.g., one person's car broke down and thus two people missed their train). The person responsible for the frustration is shown saying something, whereas the victim of the frustration is presented with a blank speech bubble.

The second type of sarcasm prompt was created by altering four preexisting *Bizarro!* comics. *Bizarro!* is a single-panel comic strip created by Dan Piraro that is syndicated daily in print newspapers across the United States. *Bizarro!* comics typically depict absurd or otherwise unlikely situations for the purpose of humor, social commentary, or both (www.bizzaro.com). The specific *Bizarro!* comics used in this study were four desert island comics, each depicting two people stranded on a small desert island in the middle of an ocean. The original cartoons all contained a single speech bubble for one of the speakers, which was made blank for the purposes of this study.

Finally, the third type of sarcasm prompt was developed using a free comic creation website (www.pixton.com). These four comics are each comprised of three panels, and each involve two speakers. In each comic, the first two panels set up an initial situation (e.g., a young man is recruited to join the army and is guaranteed to travel the world in an exciting manner by a military recruiter), while the final panel includes one of the speakers with an empty speech bubble in a situation designed to prompt a sarcastic response (e.g., the young man ends up peeling potatoes instead of traveling the world). In this manner, the sarcasm prompts were all similar in that they asked a participant to produce a sarcastic response. However, they also differed in the amount of context provided, with the desert island comics providing the least context, the black and white comics providing some context, and the three-panel comics
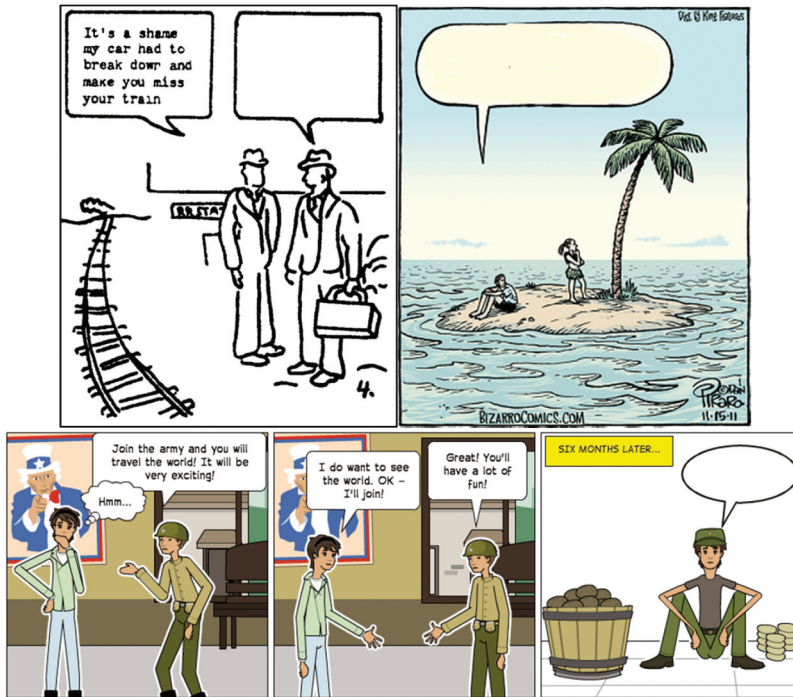
Figure 1. Examples of each type of sarcasm prompt: Black and White (top left), Desert Island (top right), and Comic (bottom).

providing the most context. An example of each type of sarcasm prompt is displayed in Figure 1, and the full list of metaphor and sarcasm prompts are included in the Supplemental Material.

Procedure

Participants first completed the Need for Cognition and Abstract Thinking surveys using a desktop computer. Afterward they began the metaphor and sarcasm production tasks. The procedure of the production tasks was verbally described to the participants, and they were asked to be as creative as possible while also aligning their answers as best they could to the definitions of metaphor and sarcasm supplied in the instructions. Participants then completed either all the metaphor prompts or all the sarcastic prompts in one block, and these blocks were randomized among participants.

For the metaphor production tasks, participants were first provided with a definition of metaphor and several examples of metaphors (none of the examples were related to the prompt stimuli). The definition of metaphor provided to participants was: *A metaphor is a comparison between two things in order to help describe something*. During each trial, participants were presented with a screen that displayed the Topic and Description in clearly marked areas of the screen, with a blank text box for the participants to type their metaphor using the keyboard. Additionally, the definition of metaphor was repeated on the screen for participants during each trial, as was a reminder of their task and the need to avoid repeating words listed in the Description. After providing answers for all 22 conventional metaphor prompts (in a random order), participants then wrote metaphors for the two novel metaphor situations, also in a random order.

For the sarcasm production tasks, participants were provided with a definition of sarcasm and given several different examples of sarcasm. The definition of sarcasm provided to participants was: *Sarcasm is a form of indirect language. When someone is being sarcastic, they mean something different than what they literally said*. Each trial randomly displayed one of the 12 sarcasm comic prompts above a text box. Participants were instructed to imagine they were the speaker depicted with the empty

speech bubble and to write something sarcastic they would say if they were in that situation. Participants' written answers for the metaphors and sarcastic replies as well as the time spent creating these answers was recorded by the computer. Participants were informed that there were no time limits for the metaphor or sarcastic reply production tasks.

Human ratings of figurative language production

An analytic rubric was created to measure participants' ability to develop creative examples of metaphor and sarcasm using two separate categories. The first category, Novelty, measured the originality of the metaphor or sarcastic reply based on other examples provided by participants as well as general familiarity of metaphors and sarcastic replies commonly used in English. The second category, Mirth, measured the emotional reaction typically associated with humor, wherein one can experience slight amusement to intense hilarity arising from humorous, interesting, or clever stimuli (Martin, 2007). Together, these categories capture aspects of creativity in general (Runco & Jaeger, 2012).

Two additional categories were also included in order to explore whether constructs related to theoretical definitions of metaphor and sarcasm would be important predictors of creativity. Accordingly, the metaphor section included the Conceptual Distance category, which measured the relative distance of conceptual domains between the entities included in each metaphorical comparison (i.e., between the noun phrase listed in the Topic and the comparative information provided by participants in each metaphor they produced). The sarcasm section of the rubric contained the Incongruity category, which measured whether the literal form of the sarcastic comment clashed in some way with the intended meaning of the comment based on the immediate context. Each category was measured using a range of 1 through 6, with a score of 1 meaning the answer did not meet the criterion in any way and a score of 6 meaning the answer met the criterion in every way. The analytic rubric is presented in full in the Supplemental Materials.

Two undergraduate assistants were recruited and trained to provide ratings of the participants' metaphor and sarcasm answers using this analytic rubric. Both assistants were undergraduate majors in linguistics in their fourth and final year. The raters first practiced using the rubric with a corpus of metaphors and sarcastic replies collected from pilot sessions using the same stimuli. After these initial ratings and discussion, the raters then scored the metaphors and sarcastic replies produced from the 61 participants in this study. For these ratings, a third rater (the author) adjudicated any disagreements between the raters of two points or greater. Any participant answers that were not examples of metaphor/ simile or sarcasm were also removed. After adjudication, the Novelty and Mirth scores were averaged to provide a single score per category per item (see below for explanation), and a total of 1304 metaphors and 716 sarcastic replies remained. Table 1 displays examples of metaphors and sarcastic replies for a range of ratings after adjudication. The full data is available in the Supplemental Materials.

Data preparation

The metaphors and sarcastic answers were analyzed for extreme outliers based on the total production times. Outliers were defined as responses greater than 2.5 standard deviations from the mean and were calculated individually for conventional and novel metaphors as well as for each of the three sarcasm prompts. This process identified 33 outliers for the metaphors (thirty conventional and two novel; 2.53% of the total data) and 21 outliers for the sarcastic replies (five Black and White, eight Desert Island, and eight three-panel comic; 2.93% of the total data). All of the following analyses were conducted with these outliers removed from the data. The final data thus included 1271 metaphors (1160 conventional, 111 novel) and 695 sarcastic responses (236 Black and White, 227 Desert Island, and 232 three-panel Comic).

Table 2 reports nonparametric correlations among the different rating categories. As can be seen, correlations between Novelty and Mirth were relatively strong for both metaphors and sarcastic

Table 1. Examples of metaphors and sarcastic replies with different ratings.

Type	Example	Rating
Conventional Metaphor	*Some property is worth a fortune.*	1
	Some jobs look like being in jail.	2
	Some old men are like trees that can never be cut down.	3
	The Earth is like a subway in Tokyo.	4
	Kinship can bring family together like experiments with the Mythbusters.	5
Novel Metaphor	*Being in that class was like watching the grass grow.*	2
	Being in that class was like staring at a brick wall for 90 minutes.	3
	Eating fried chiken feet was like licking somebody's dirty foot.	4
	Being in that class was like rolling a rock up a hill every day, only to watch it fall back down again.	5
Black and White	*I can see the back of her head.* [blocked movie view]	2
	I am so glad I could help make it look better! [broken vase]	3
	It's fine these are just my interview clothes. [puddle splash]	4
	Never mind, I never watch a train leaving before. Isn't it beautiful? [missed train]	5
Desert Island	*We're finally on that vacation you wanted*	2
	You look like you're having fun over there.	3
	The sunsets make this all worth it!	4
	What do you want to do today? I was thinking we could get my friends together and play a game of volleyball.	5
Three-Panel Comic	*Great timing Sheryl … * [late classmate]	2
	*Look at me, defying the odds … * [race braggart]	3
	I can transfer schools if you still want to be the fastest runner here. [race braggart]	4
	Wow! I didn't know the world looked like the inside of a kitchen. [joining army]	5

Typographical errors made by participants are retained. Context provided in brackets [] for examples in the Black and White and Three-Panel Comic categories.

Table 2. Correlations among analytic rubric rating categories.

Metaphors	Novelty	Mirth
Conceptual Distance	0.248	0.255
Novelty		0.825
Sarcastic Replies	*Novelty*	*Mirth*
Incongruity	−0.045	0.068
Novelty		0.741

Correlations calculated using the Kendall's tau method and are thus more conservative compared to Pearson's *r*.

replies, suggesting they were measuring similar constructs. Further attempts at dimension reduction using principal components analysis failed to yield meaningfully distinct components between Novelty and Mirth. Therefore, these measures were averaged per item to create a single dependent variable for each figurative language type. The Conceptual Distance and Incongruity scores were kept separate to be used as predictor variables in the statistical analysis.

Statistical analysis

Two linear mixed-effects regression models were fit in order to test whether Need for Cognition, Abstract Thinking, production time, native English speaker status, metaphor Conceptual Distance, or sarcasm Incongruity scores were predictive of the novelty/mirth ratings for the metaphors and sarcastic replies. Each model included the averaged metaphor or sarcasm novelty/mirth ratings as the dependent variable with participants and items entered as crossed random effects. For the metaphor model, metaphor type was entered as a dummy-coded categorical variable (conventional or novel) and interactions were fit between metaphor type and all the remaining predictor variables: metaphor production time, metaphor Conceptual Distance ratings, native English speaker status (yes/ no), and participants' Abstract Thinking and Need for Cognition scores. For the model testing sarcasm ratings, sarcasm prompt was entered as a dummy-coded categorical variable (comic, black and white, or desert island) and interactions were fit between prompt and the other predictor variables: sarcasm

production time, sarcasm Incongruity ratings, native English speaker status (yes/no), and participants' Abstract Thinking and Need for Cognition scores. Significance and inclusion of effects was assessed using relative log-likelihood comparisons with the Akaike's Information Criterion (AIC) as a criterion of model fit (Tremblay & Ransijn, 2015). Any terms included in significant interactions were also modeled as separate main effects in order to properly interpret their significance as simple main effects, and random slopes were tested for metaphor type and sarcasm prompt on the random intercept of subject. Model effect size (R^2) was obtained using the *MuMIn* package, which provides a marginal (fixed effects only) and conditional (fixed and random effects) R^2 value.

Results

Survey data

Table 3 includes descriptive statistics for participant answers to the Need for Cognition (NFC) and Abstract Thinking (AT) surveys and internal consistency measures. The results from Chronbach's alpha reports good internal consistency for both surveys.

Metaphor production

Table 4 displays the mean and standard deviation of the averaged Novelty/Mirth ratings, Conceptual Distance ratings, and production times for the metaphors, categorized by metaphor type. What cannot be seen in Table 4 is that no metaphors were given a rating above 5 for Conceptual Distance or for the Novelty/Mirth scores. As can be seen, the average production times and Novelty/Mirth ratings were higher for the novel metaphors when compared to the conventional metaphors. Little difference is seen in Conceptual Distance ratings, which are on average close to the effective maximum rating of 5, suggesting a ceiling effect for this measure.

In order to better understand the numerical information in Table 4, Figure 2 displays violin plots for the production times and Novelty/Mirth ratings of the novel and conventional metaphors. These plots suggest the production times for conventional metaphors were skewed toward shorter production times, while the production times for the conventional metaphors were more uniform yet generally took more time to create. For the Novelty/Mirth ratings, these plots reveal that the lowest Novelty/Mirth rating for the novel metaphors was 2, with a relatively stable distribution for ratings of 2 through 5. Conversely, the conventional metaphors received ratings from 1 to 5 but were also skewed toward an uneven distribution of Novelty/Mirth scores, with a higher frequency of scores of 2 when compared to the other rating options. Because the Conceptual Distance ratings were heavily skewed

Table 3. Descriptive statistics of survey results.

Need for Cognition					
M	*SD*	*Max*	*Min*	*Range*	*Chronbach's α*
19.11	17.03	64	−11	75	.84
Abstract Thinking					
M	*SD*	*Max*	*Min*	*Range*	*Chronbach's α*
13.69	5.09	25	3	22	.81

Table 4. Descriptive statistics for metaphors.

		Production Time		Novelty/Mirth Rating		Conceptual Distance	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Conventional	1160	50.296	34.465	3.255	1.178	4.553	0.892
Novel	111	76.372	41.832	3.687	1.030	4.649	0.712
All	1271	52.574	35.914	3.293	1.172	4.561	0.878

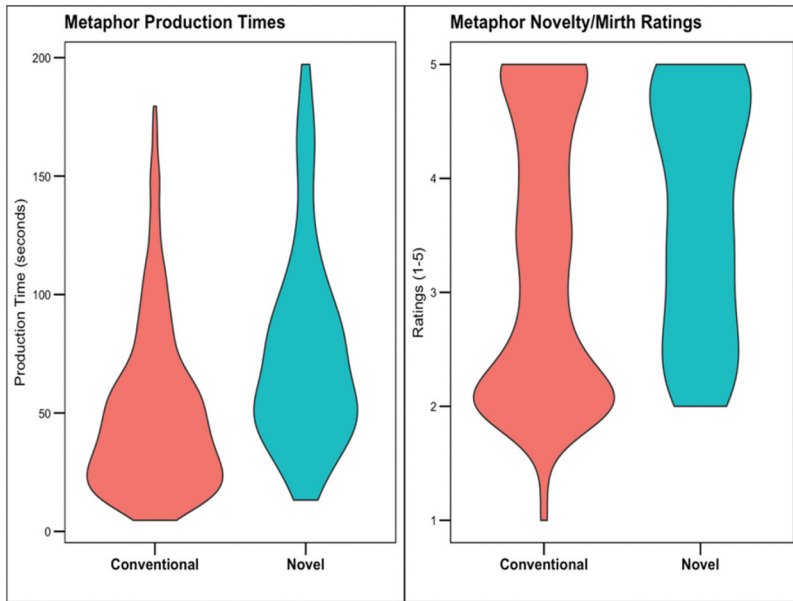Production time is measured in seconds.

Figure 2. Violin plots depicting distribution of production time (left) and novelty/mirth ratings (right) for produced metaphors.

toward the top end of the scale, violin plots did not reveal any insight into the distribution of these scores. Figure 3 instead displays these scores using bar plots with the Conceptual Distance ratings sorted into two groups: ratings between 1 and 4 and ratings of 5. Approximately 70% of both the conventional and novel metaphors received Conceptual Distance ratings of 5, providing further evidence of a ceiling effect for these ratings.
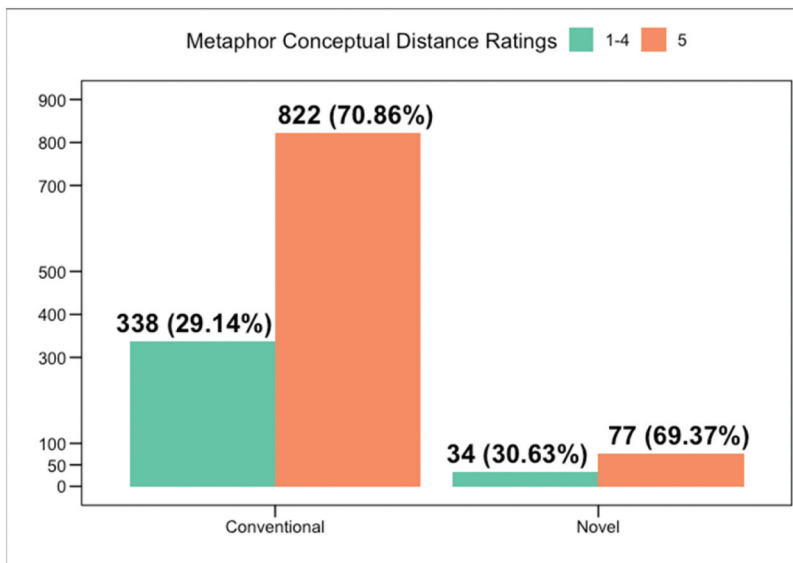

Figure 3. Bar plot depicting distribution of metaphor Conceptual Distance ratings for conventional and novel metaphors. For ease of interpretation and to better understand the distribution of the data, Conceptual Distance ratings have been grouped into scores between 1 and 4 and scores of 5. Note that the lowest Conceptual Distance rating for novel metaphors was 2.

Predicting metaphor novelty/mirth scores

The regression model predicting metaphor Novelty/Mirth ratings included significant main effects for production time and Conceptual Distance ratings. Both of these effects were positively associated with the Novelty/Mirth ratings in that increases in production time and conceptual distance ratings both predicted higher Novelty/Mirth ratings. The model also included a significant interaction between metaphor type and Need for Cognition. This interaction indicated that higher levels of Need for Cognition were associated with significantly higher Novelty/Mirth ratings for novel metaphors when compared to conventional metaphors (see Figure 4). A random slope of this effect produced a singular fit, suggesting it did not explain any additional variance and was thus left out of the final model. The marginal R^2 for this model was .174 and the conditional R^2 was .348. Table 5 displays the coefficients, standard errors, t and p values for the terms included in the final model.
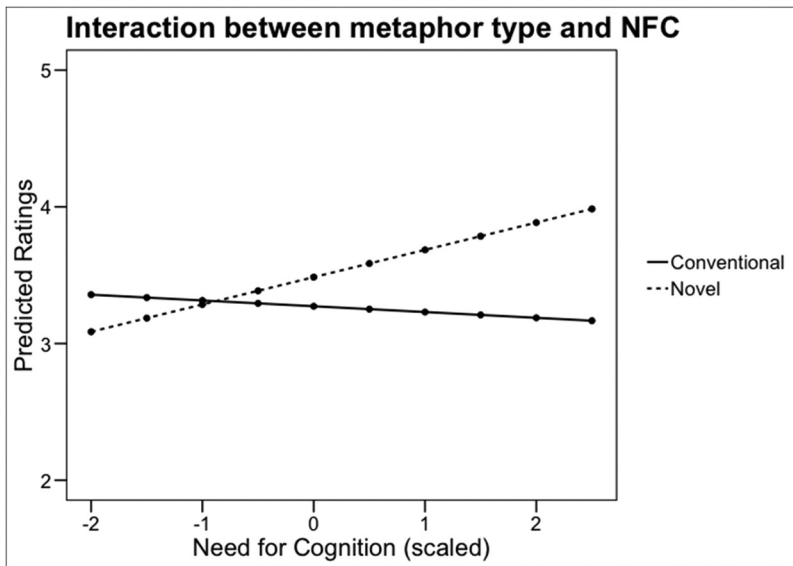

Figure 4. Interaction between metaphor type and Need for Cognition (NFC). Higher NFC values were associated with significantly higher Novelty/Mirth ratings for metaphors produced in response to the novel metaphor prompts when compared to the conventional metaphor prompts.

Table 5. Linear mixed effects model predicting metaphor novelty/mirth ratings.

Random	Variance	SD	Marginal R^2	Conditional R^2
Subjects	0.144	0.380	0.174	0.348
Items	0.090	0.300		
Fixed	*Estimate*	*SE*	*t*	*p*
(Intercept)	1.264	0.173	7.285	<.001
Production Time	0.246	0.030	8.233	<.001
Type: Novel	0.213	0.241	0.881	0.388
Need for Cognition	−0.042	0.057	−0.743	0.461
Conceptual Distance Rating	0.441	0.033	13.222	<.001
Type: Novel * Need for Cognition	0.242	0.095	2.551	0.011

Type baseline = Conventional.

Sarcasm production

Table 6 displays the mean and standard deviation for the averaged Novelty/Mirth ratings, production times, and Incongruity ratings for the sarcastic replies, categorized by the three different sarcasm prompts. As can be seen, the Incongruity ratings are on average relatively high and represent a ceiling effect with little variation among the three prompts, similar to the results obtained for the Conceptual Distance ratings of the metaphors. Also similar to the metaphor ratings, no sarcastic replies received Incongruity or Novelty/Mirth ratings higher than 5. The results for the production times suggest some consistency in terms of how long participants spent writing their sarcastic replies for the three different prompts. Finally, some variation is suggested among the three prompt types for the Novelty/Mirth ratings, in that sarcastic replies made in response to the comic prompts have descriptively lower Novelty/Mirth ratings compared to the other prompts.

These observations are further supported when considering the same data depicted using violin plots (see Figure 5). Distribution of the production times suggests a relatively similar

Table 6. Descriptive statistics for sarcastic replies.

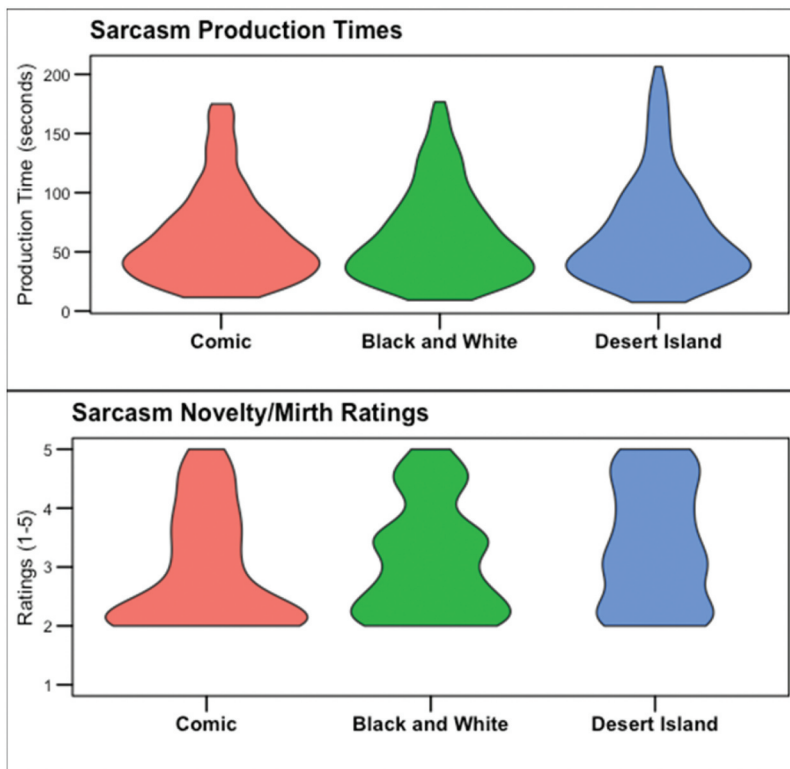|		Production Time		Novelty/Mirth		Incongruity	
Sarcasm Prompt	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Comic	232	63.656	39.286	2.947	0.968	4.420	0.924
Black and White	236	63.752	38.624	3.118	0.929	4.612	0.729
Desert Island	227	66.923	43.278	3.366	1.010	4.564	0.733
All	695	64.756	40.387	3.142	0.982	4.532	0.803

Production time is measured in seconds.

Figure 5. Violin plots depicting distribution of production time (top) and Novelty/Mirth ratings (bottom) for produced sarcastic replies.

distribution for all three sarcasm prompts, with a general skew toward times under one minute. The distribution of the Novelty/Mirth ratings suggests a skew toward lower ratings for sarcastic replies produced in response to the Three-Panel Comic prompts, a more tapered upward skew for the Black and White prompts, and relatively uniform distribution for the Desert Island prompts. Notably, the violin plots reveal that none of the sarcastic replies in these data received a Novelty/Mirth rating lower than 2. Similar to the distribution of Conceptual Distance ratings for the metaphors, the data in Table 6 suggest that Incongruity ratings for the sarcastic replies were skewed toward higher ratings. This data is plotted as a bar plot in Figure 6, which groups the Incongruity ratings into categories between 2 and 4 and ratings of 5 and provides further evidence of this skew.

Predicting sarcastic reply novelty/mirth scores

The regression model predicting the averaged Novelty/Mirth ratings for sarcastic replies included significant main effects for production time, Incongruity ratings, and sarcasm prompt. The production times and Incongruity ratings were both positively associated with the Novelty/Mirth ratings, in that increases in production time and Incongruity ratings both predicted higher Novelty/Mirth ratings for the sarcastic replies. Additionally, sarcastic replies made in response to the Desert Island prompts were predicted to have significantly higher Novelty/Mirth ratings when compared to replies made in response to the Three-Panel Comic prompts. There were no significant differences between the Black and White and three-panel Comic prompts. Subsequent pairwise comparisons using the estimated marginal means reported no significant differences between the Black and White and Desert Island prompts. These differences are summarized in Figure 7. There were no significant interactions, and a random slope of prompt type fit on the random slopes of subjects resulted in a singular fit and was thus left out from the final model. The marginal R^2 for this model was .064 and the conditional R^2 was .251. Table 7 displays the coefficients, standard errors, t and p values for the terms included in the final model.
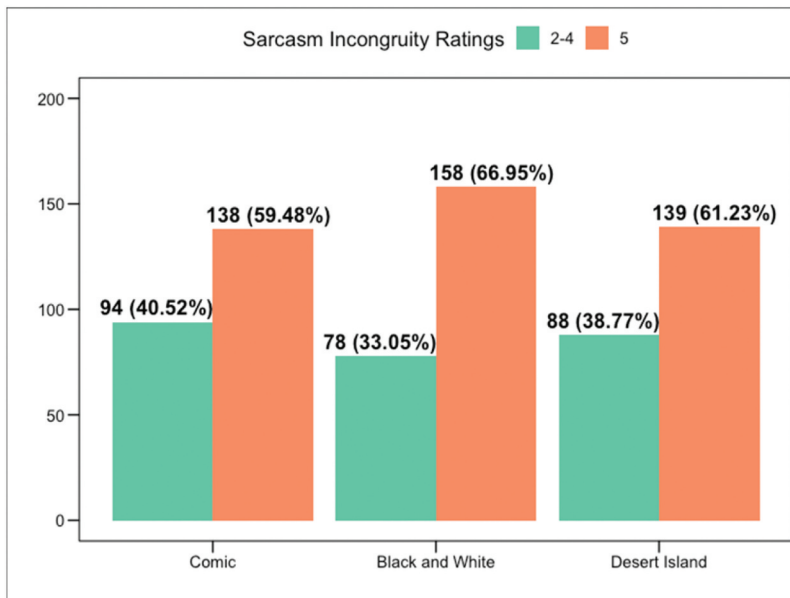

Figure 6. Bar plot depicting distribution of sarcastic reply Incongruity ratings among different sarcasm prompts. For ease of interpretation, Incongruity ratings have been grouped into ratings between 2 and 4 and ratings of 5 (no sarcastic reply in this data received an Incongruity rating lower than 2 or higher than 5).
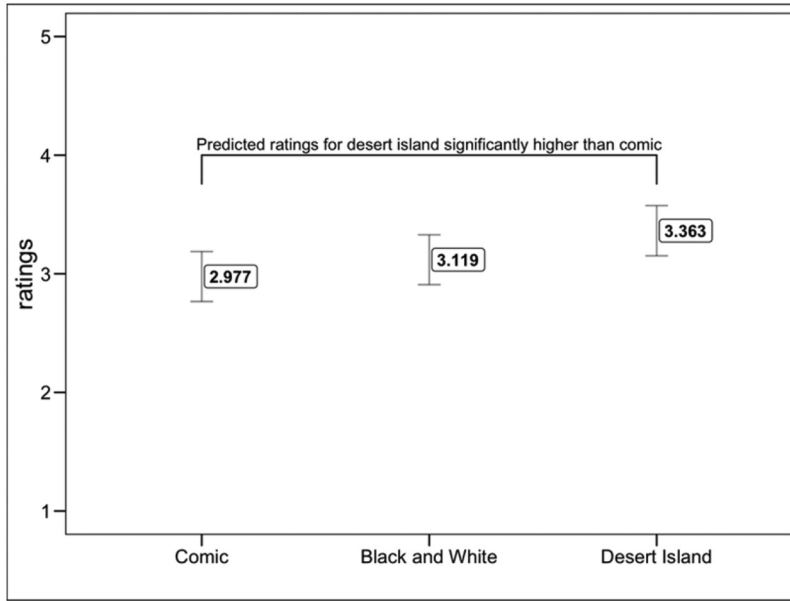
Figure 7. Sarcastic replies made in response to Desert Island prompts were predicted to have significantly higher Novelty/Mirth ratings when compared to sarcastic replies made in response to three-panel Comic prompts ($p = .018$). There were no other significant differences among the prompt categories.

Table 7. Linear mixed effects model predicting sarcastic reply novelty/mirth ratings.

Random	Variance	SD	Marginal R^2	Conditional R^2
Subjects	0.159	0.398	0.064	0.251
Items	0.023	0.151		
Fixed	*Estimate*	*SE*	*t*	*p*
(Intercept)	2.126	0.236	9.027	<.001
Production Time	0.004	0.001	3.965	<.001
Prompt: Black and White	0.141	0.133	1.060	0.317
Prompt: Desert Island	0.386	0.134	2.885	0.018
Incongruity Ratings	0.134	0.045	2.965	0.003

Prompt baseline = Three panel Comic.

Discussion

The purpose of this study was to further explore associations among individual differences and elicited figurative language production. Participants produced either metaphorical or sarcastic comments in response to different prompts, and these comments were rated for novelty, mirth and either conceptual distance (metaphors) or incongruity (sarcastic replies) by trained human raters. Participants' levels of Abstract Thinking (AT) and Need for Cognition (NFC) were also measured using preexisting surveys. These features, along with production time, prompt type, and native English speaker status were used to predict the novelty and mirth ratings for the metaphors and sarcastic responses.

Novelty and mirth ratings

The results from the human ratings indicated a strong overlap between perceptions of novelty and perceptions of mirth, supported by significant, strong positive correlations. While the novelty category was straightforward in that it measured the relative originality of a participants' answer, mirth was a catch-all category designed to measure any aspects of the participant answers that the human raters found interesting, clever, or humorous. Because the novelty and mirth scores were so strongly

correlated for both metaphors and sarcastic responses, it may be that it was difficult for the human raters to disassociate perceptions of novelty and mirth, suggesting that these two constructs are strongly associated with one another. In other words, in these data, the metaphors and sarcastic responses that were humorous, clever, or interesting were also those that were original. In hindsight it does not seem surprising that original or novel examples of metaphor or sarcasm would also be perceived as mirthful or humorous, but to obtain such a strong overlap between these two constructs was surprising because not all creative products are automatically humorous, and not all examples of humor are creative. Therefore, the results from these ratings may more strongly reflect aspects of creativity unique to metaphorical comparisons and sarcastic responses, with which humor is strongly associated (Attardo, 2017; Veale, 2013). As such, disentangling the differences between these two constructs in the context of figurative language production would be an important next step for research in this area.

Conceptual distance and incongruity ratings

There was an observed ceiling effect for Conceptual Distance and Incongruity ratings in these data. This means that most of the metaphors were perceived to involve conceptually distant entities, and that most of the sarcastic replies were perceived to be incongruous with the context in which they were spoken. This may have been a result of the instructions provided to participants, who were informed to match their examples of metaphor or sarcasm to the definitions provided while also being as creative as possible. Regardless, for both metaphors and sarcastic replies, human ratings for the theoretical constructs associated with metaphors or sarcastic replies were significant, positive predictors of the novelty/mirth ratings. Thus, while having high ratings of these elements are not necessary in order for them to fit the definitions of metaphors or sarcastic replies, they may yet influence perceptions of originality or humor associated with them.

Production times

Longer production times were associated with higher perceptions of novelty/mirth for both the metaphors and the sarcastic replies. This replicates findings from one prior elicited metaphor production study which also found that participants with longer production times created metaphors which were more highly rated (Silvia & Beaty, 2012). Although Pierce and Chiappe (2008) reported a different trend, with quality of metaphorical vehicles declining as participants spent more time creating them, this difference is likely explained by variation in the types of metaphors participants were asked to produce. In Pierce and Chiappe (2008), participants were responding to prompts more likely to evoke conventionalized metaphors, whereas Silvia and Beaty (2012) had participants generate novel metaphors with the same prompts used in the current study. For the current study, participants were not only asked to produce both conventional and novel metaphors, they were also instructed to be as creative as possible and that they were free to spend as much time as necessary. As such, when it comes to creative metaphor generation, these results in tandem with those from Silvia and Beaty (2012) cohere to suggest that, when given sufficient time, participants can generate metaphors that are perceived to be more unique and mirthful. By extension, participants who spend less time are likely drawing from previously encountered (and thus less novel) metaphors, whereas those who spend more time are likely engaging in more purposeful cognitive behavior to select appropriate and creative entities for their metaphorical comparisons (Silvia & Beaty, 2012). A similar effect was also found in the current study for the sarcastic replies and is likely explained in the same manner. When generating a sarcastic reply, participants had more time to consider the contextual situation depicted in the prompt and generate a reply that fit with the definition of sarcasm they were given while also being creative.

Native English speaker status

In the current data, no differences were observed for whether a participant was a native or non-native speaker of English. While this variable was mainly included to control for potential differences between these two categories of speakers, this finding is somewhat surprising in light of the prior vocabulary

effects reported for metaphor production quality. The international student participants in this study were required to be approximately at a CEFR level of B2 or higher in order to enter the university. However, even at advanced proficiency levels, non-native English speakers will still lack the depth and breadth of vocabulary knowledge that native speakers possess, but this difference did not appear to influence the perceptions of novelty/mirth among the raters in these data. A likely explanation for the lack of a difference between native and non-native English speakers is that there were no time restrictions imposed on the participants during the metaphor or sarcastic reply production tasks. The other two production studies which have reported vocabulary effects for metaphor production required participants to complete the metaphor generation task in 15 minutes (Beaty & Silvia, 2013; Chiappe & Chiappe, 2007). As the current task was untimed, the non-native English-speaking participants likely had sufficient time to draw upon their existing linguistic resources and experience living in the United States to craft their metaphors and sarcastic replies. In all, this finding suggests that the data between native and non-native speakers were comparable in the current study, due in part to the nature of the production task.

Need for cognition and abstract thinking

Need for Cognition (NFC) was a significant predictor of novelty/mirth ratings for the metaphors but not for the sarcastic replies. Moreover, NFC interacted with metaphor prompt type in that higher levels of NFC were associated with significantly higher novelty/mirth ratings for novel metaphor prompts when compared to the conventional metaphor prompts. In other words, this finding suggests that participants who indicated a predilection toward engaging in cognitively demanding activities were predicted to produce novel metaphors perceived as more creative by the human raters. In contrast, higher NFC had relatively little effect on perceptions of novelty/mirth associated with the conventional metaphors. In this manner, the results from NFC here align with the findings from Beaty and Silvia (2013), who reported measures of intelligence representing problem-solving ability were associated more strongly with perceptions of quality in novel metaphors when compared to conventional metaphors.

As such, this study provides additional evidence that individual differences in cognitive resources and preferences interact with the type of metaphorical comparisons one is asked to create. It could also be argued that this finding may simply reflect a difference between a free and constrained production task, which may not be specifically related to metaphor production. However, there were no significant results for NFC and the sarcastic replies, which also included prompts of varying freedom (see below). Therefore, this study, in combination with the findings of prior studies provides further evidence suggesting that producing creative metaphors can place specific, measurable demands on cognitive systems. Participants who possess more problem-solving ability and who enjoy cognitive challenges are more likely to produce metaphors that are perceived as novel and mirthful when given complete freedom to do.

There were no effects found for Abstract Thinking (AT) for the metaphors or the sarcastic responses. This suggests that whether one tends to conceptualize events as abstract or concrete makes little difference in regard to perceptions of novelty/mirth associated with metaphors or sarcastic replies. Huang et al. (2015) described sarcasm as a potential catalyst for creativity because it was observed to increase AT in their participants. However, Huang et al. (2015) also instructed participants to consider the first sarcastic comment that came to mind but did not further analyze participants' sarcasm production. Perhaps AT is thus only important when participants must solve problems (in this case, saying something sarcastic) as quickly as possible. Another consideration is that the relationship between AT and sarcasm may be unidirectional, in that sarcasm use fosters levels of AT (and subsequently creativity) but that AT does not necessarily influence the ability to generate sarcastic replies. Future research exploring AT in different production contexts is thus warranted.

Sarcasm prompt type

There was a significant difference among the three sarcasm prompts in that novelty/mirth ratings for sarcastic replies made in response to the Three-Panel Comic prompts were predicted to be

significantly lower when compared to sarcastic replies made in response to the Desert Island prompts. This finding is likely related to the differences in context available to the participant when making the sarcastic replies. In the Three-Panel comic prompts, specific situations were presented to the participant which may have constrained the range of sarcastic replies available. For instance, the Three-Panel Comic example in Figure 1 depicts a new military recruit sitting alone and peeling potatoes after being promised to travel the world. Many of the sarcastic replies for this comic drew upon the contrast between the promised excitement of joining the military and the relative mundaneness of potatoe peeling. As such, responses such as *Ah, the world of potatoes!* and *He was right this was fun and exciting* suggest it was relatively simple for participants to create sarcasm by echoing the recruiter's promise and alluding to failed expectations (Kumon-Nakamura, Glucksberg, & Brown, 1995; Wilson & Sperber, 1992). The other comic prompts all constructed similarly constrained contexts, such as a man bragging about being a fast runner before subsequently losing a race, a weather reporter promising good weather only to have it rain, and a classmate promising to help work on an assignment but not arriving until a few minutes before the assignment is due. In the same manner as the example in Figure 1, participants in this study were able to repeat and echo early parts of the discourse to create a sarcastic comment supported by the context in revealed by the end of the comic.

In contrast, the Desert Island prompts provided no prior context and no prior discourse for participants to rely upon. All four of the Desert Island comics included two people, with one speaking to the other, next to a palm tree in a single panel. Aside from some minor differences in the behavior of the people in each comic (e.g., one comic shows a man writing a note while another shows a woman staring at the palm tree), these prompts are almost identical in terms of the sparse context provided. Without any prior conversation or discourse to echo (or knowledge of the relationship between the people in the comic), participants were forced to fill the contextual gaps with their own creative hypothetical information based on the immediate moment in time depicted by each comic. As such, the same entities depicted in the comics were portrayed by the participants in different ways. For example, the prompt with a man writing a note received responses such as *I'm sure writing your memories will help us get out of here* and *I hope you are drawing my good profile*. The end result was that these prompts allowed for more variation in the types of answers provided when compared to the Three-Panel comic prompts, which may explain the significant differences in perceptions of novelty/mirth made by the raters.

The Black and White comics were likely somewhere between the two other prompt types in terms of the amount of prior context and discourse provided to the participants. Like the Desert Island comics, the Black and White comics only contained a single panel. At the same time, the Black and White comics also included an additional speaker to respond to as well as a clearly depicted frustrating situation (i.e., not being able to see at a movie theater, being splashed by a car, breaking someone's vase, and missing a train). This explains why the Black and White prompts did not differ significantly from either of the other two sarcasm prompt types. As a whole, these results point to the importance of prompt information when conducting elicited sarcasm production research studies, in that the amount of context provided in a sarcasm prompt may seriously influence the originality of the response provided.

Conclusion and future directions

This study provides new insight from elicited figurative language production studies, which are relatively rare in the figurative language research literature. One particularly challenging aspect relates to how figurative language ability should be measured. Based on prior research and connections between figurative language and creativity, the approach taken in this study was to measure perceptions of figurative language associated with aspects of creativity. However, there are likely other ways to measure figurative language ability in relation to metaphor and sarcasm and these can be explored in future elicited figurative language production studies. Another important implication from the results of this study demonstrate the effects of task and prompt characteristics in terms of the types of

metaphors and sarcastic replies generated. In general, it seems that prompts which place too many constraints on the context and language will result in less original or creative examples of figurative language production. Finally, while there were no significant effects found for Abstract Thinking, this study provides yet another link between Need for Cognition and figurative language use, suggesting that it is an important individual difference to consider in future figurative language production research.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Stephen Skalicky http://orcid.org/0000-0001-8329-4531

References

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, *32*(6), 793–826. doi:10.1016/S0378-2166(99)00070-3

Attardo, S. (Ed.). (2017). *The Routledge handbook of language and humor*. New York, NY: Routledge.

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory & Cognition*, *41*(2), 255–267. doi:10.3758/s13421-012-0258-5

Cacioppo, J. T., Petty, R., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. doi:10.1207/s15327752jpa4803_13

Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, *56*(2), 172–188. doi:10.1016/j.jml.2006.11.006

Colston, H. L. (2015). *Using figurative language*. New York, NY: Cambridge University Press.

Colston, H. L. (2017). Irony and sarcasm. In S. Attardo (Ed.), *The Routledge handbook of language and humor* (pp. 234–249). New York, NY: Routledge.

Corts, D. P., & Meyers, K. (2002). Conceptual clusters in figurative language production. *Journal of Psycholinguistic Research*, *31*(4), 391–408. doi:10.1023/A:1019521809019

Deignan, A., Littlemore, J., & Semino, E. (2013). *Figurative language, genre and register*. Cambridge, MA: Cambridge University Press.

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 811–828.

Gerrig, R. J., & Gibbs, R. W. (1988). Beyond the lexicon: Creativity in language production. *Metaphor and Symbol*, *3*(3), 1–19.

Gibbs, R. W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*(1), 3.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge, MA: Cambridge University Press.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*(1–2), 5–27.

Gibbs, R. W., & Colston, H. L. (Eds.). (2007). *Irony in language and thought: A cognitive science reader*. New York, NY: Lawrence Erlbaum Associates.

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*. New York, NY: Cambridge University Press.

Giora, R. (2003). *On our mind: Salience, context, and figurative language*. New York, NY: Oxford University Press.

Haught, C. (2013). A tale of two tropes: How metaphor and simile differ. *Metaphor and Symbol*, *28*(4), 254–274.

Huang, L., Gino, F., & Galinsky, A. D. (2015). The highest form of intelligence: Sarcasm increases creativity for both expressers and recipients. *Organizational Behavior and Human Decision Processes*, *131*, 162–177.

Hussey, K. A., & Katz, A. N. (2006). Metaphor production in online conversation: Gender and friendship status. *Discourse Processes*, *42*(1), 75–98.

Ivanko, S. L., Pexman, P. M., & Olineck, K. M. (2004). How sarcastic are you?: Individual differences and verbal irony. *Journal of Language and Social Psychology*, *23*(3), 244–271.

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes*, *51*(4), 287–311.

Katz, A. N., & Pexman, P. M. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. *Metaphor and Symbol*, *12*(1), 19–41.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*(1), 3.

Low, G., Littlemore, J., & Koester, A. (2007). Metaphor use in three UK University lectures. *Applied Linguistics*, *29*(3), 428–455.

Martin, R. A. (2007). *The psychology of humor: An integrative approach*. Burlington, MA: Elsevier Academic Press.

Nacey, S. (2013). *Metaphors in learner English*. Amsterdam, Netherlands: John Benjamins Publishing.

North, B., Ortega, A., & Sheehan, S. (2011). *British Countil—EAQUALS Core Inventory for General English*. London, UK: British Council/European Association for Quality Language Services.

Olkoniemi, H., Johander, E., & Kaakinen, J. K. (2018). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*. doi:10.3758/s13421-018-0852-2

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433–450.

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

Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science*, *5*(3), 159–163.

Rosenzweig, S. (1945). The picture-association method and its application in a study of reactions to frustration. *Journal of Personality*, *14*(1), 3–23.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96.

Silvia, P. J., & Beaty, R. E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence*, *40*(4), 343–351.

Skalicky, S. (2019). Investigating satirical discourse processing and comprehension: The role of cognitive, demographic, and pragmatic features. *Language and Cognition*, *11*, 499–525.

Tremblay, A., & Ransijn, J. (2015). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. *R Package Version 2.10*. Retrieved from https://CRAN.R-project.org/package=LMERConvenienceFunctions

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, *57*(4), 660.

Veale, T. (2012). *Exploding the creativity myth: The computational foundations of linguistic creativity*. New York, NY: Bloomsbury.

Veale, T. (2013). Humorous similes. *Humor*, *26*(1), 3–22.

Whalen, J. M., Pexman, P. M., & Gill, A. J. (2009). "Should be fun—Not!": Incidence and marking of nonliteral language in e-mail. *Journal of Language and Social Psychology*, *28*(3), 263–280.

Wilson, D., & Sperber, D. (1992). On verbal irony. *Lingua*, *87*(1), 53–76.