

# Music-induced Positive Mood Stimulates Metaphor Production

Laura Pissani<sup>1</sup>, Magdalena-Victoria Meiser<sup>1</sup>, and Vera Demberg<sup>1,2</sup>

{lpissani, mvmeiser, vera}@lst.uni-saarland.de

<sup>1</sup>Department of Language Science and Technology

<sup>2</sup>Department of Computer Science

Saarland University

Saarbrücken, Germany

## Abstract

Metaphors are a creative use of language that conveys complex ideas through abstract reasoning and cognitive flexibility. While prior research has demonstrated that music influences creativity, its specific impact on metaphor production remains unexplored. In this study, 90 adults were assigned to one of three groups—silence, happy music, and sad music—and completed a metaphor production task, generating expressions for emotions (e.g., being happy) and actions (e.g., telling a lie). Participants also completed convergent and divergent thinking assessments to account for individual differences in creativity. Results showed that participants who listened to happy music while doing the task were more likely to produce figurative expressions, with convergent creativity positively predicting their production, while divergent creativity had no effect. Moreover, metaphors produced with background music were generally rated as more novel than those produced in silence, with sad music leading to metaphors with a more negative emotional tone. These findings suggest that extrinsic factors, particularly happy music, can enhance our ability to produce metaphors by boosting the cognitive flexibility required for creative thinking.

**Keywords:** metaphor production; creativity; mood; music

## Introduction

Metaphors are creative linguistic devices that facilitate the expression of rich and complex ideas. The ability to generate metaphors, especially novel ones, reflects higher-order cognitive processes, such as abstract reasoning and cognitive flexibility. Previous studies have identified various cognitive functions—such as working memory (Chiappe & Chiappe, 2007), selective attention and divergent thinking (Menashe et al., 2020), general reasoning and verbal fluency (Beaty & Silvia, 2013)—as well as age (Levorato & Cacciari, 2002) as factors influencing metaphor production.

Menashe et al. (2020) examined the role of several cognitive measures in producing novel and conventional similes. Participants created expressions for emotions (e.g., feeling worthless), which were classified as literal or figurative and rated for originality. They also completed assessments of vocabulary, cognitive control, selective attention, executive functioning, working memory, and divergent thinking. The results showed that selective attention and cognitive control supported the production of both novel and conventional responses, while novel responses were additionally linked to executive functioning and divergent thinking. These results suggest that generating novel expressions places greater demands on cognitive resources compared to conventional ones.

Moreover, Levorato and Cacciari (2002) found that the ability to generate metaphors increased from childhood to adolescence, with a slight decline into adulthood. However, metaphors produced by adults were often rated as more comprehensible, appropriate, and novel. These findings indicate that metaphor production improves with age as cognitive abilities such as abstract reasoning and flexibility mature.

Earlier research has demonstrated that mood can influence creativity (e.g., Adaman & Blaney, 1995; Isen et al., 1987). For instance, Callaghan and Growney (2013) showed that video-induced mood affected creativity. Participants who watched a sad video followed by happy music performed less creatively on a divergent thinking task, suggesting that mood congruency—particularly in negative emotional states—may constrain creativity. Similarly, Ritter and Ferguson (2017) examined the role of music-induced mood on creativity. In their study, participants listened to music designed to induce various moods (happiness, sadness, calmness, or anxiety) while performing divergent and convergent thinking tasks. Results indicated that happy music enhanced divergent thinking, but had no effect on convergent thinking tasks, which require a single correct solution. These findings suggest that a positive mood may enhance cognitive abilities such as fluency and flexibility, relevant for the generation of new ideas.

Despite the well-established link between mood and creativity, the specific role of mood in metaphor production remains underexplored. Metaphor generation, a form of creative thinking, involves making novel connections between seemingly unrelated concepts. Given that a positive mood enhances divergent thinking—the ability to generate original ideas—it is plausible that mood, manipulated through different musical backgrounds, could also influence metaphor production, potentially increasing both the quantity and originality of metaphorical responses.

The current study investigates whether listening to specific types of music influences the production of figurative language in adults. We employed a metaphor production task, where participants were asked to create new expressions to represent emotions (e.g., being happy) and actions (e.g., revealing a secret). Participants completed the task either in silence, with happy music, or with sad music playing in the background. We also collected individual measures of convergent and divergent creativity to ensure that any observed effects were specifically attributed to music listening.

Based on findings by Ritter and Ferguson (2017), we predicted that participants listening to happy music would generate more figurative responses than those in silence, and that metaphors produced in either music group would be perceived as more novel. Additionally, we predicted that metaphorical responses would be more frequent and judged as more appropriate for expressing emotions rather than actions, as metaphors are more commonly used to describe abstract topics than concrete ones (Al-Azary & Buchanan, 2017).

## Methodology

### Participants

A total of 90 native English speakers (age range: 19–72 years; mean age = 32.13 years, SD = 10.29; 52 female) were recruited via Prolific and completed all subtests. Participants registered blindly to one version of the study, with 30 participants recruited per group<sup>1</sup>. The study lasted approximately 30 minutes, and participants received £5 compensation.

### Materials

**Remote Associates Test** The Remote Associates Test (RAT; Mednick, 1962) measures convergent thinking, the ability to identify a single solution word that connects three given cue words. Each cue word is associated to the solution word in a different way. For example, in the set *cottage*, *swiss*, and *cake*, the first two words are semantically linked to the solution word (*cottage cheese*, *swiss cheese*), while the third cue forms a compound word (*cheesecake*). In order to determine the correct solution, individuals must creatively identify the connections among the cue words.

We selected 30 items from Bowden and Jung-Beeman (2003), who normed 144 items according to their solution rate and response time. The 30 items we selected had an average solution rate of 60% and an average response time of 6 seconds. We implemented a 15-second time limit for each item, with a beep sound played 3 seconds before time expired.

**Divergent Association Task** The Divergent Association Task (DAT; Olson et al., 2021) measures divergent thinking, the ability to generate multiple solutions to open-ended problems. Participants have to name ten words that are as unrelated as possible. The diversity of responses is quantified using semantic distance, calculated based on the unrelatedness of the meanings of the provided words. Higher semantic distance scores indicate greater divergent thinking ability. This test has been shown to correlate with the Alternative Uses Task (AUT; Guilford, 1950), in which participants generate alternative uses for common objects, judged on originality, flexibility, and fluency.

**Metaphor Production Task** We employed a metaphor production task using emotion and action prompts from Levorato and Cacciari (2002). Participants were presented with 18 prompts, one at a time (see Table 1), and instructed to write

a different way to express the human experience. To prevent responses that included qualifiers (e.g., good, bad, normal) or nonsensical answers, we provided clear instructions with examples. Participants were asked to reflect on the emotions associated with each prompt and recall an experience that evokes similar emotions. For example, if the prompt was “Being alone is,” participants may associate it with positive or negative experiences and reflect that in their answers (e.g., *charging my social battery*, *drowning in the ocean*). This task included three versions: no background music, happy background music, and sad background music. In the music versions, 15 seconds of music were played on a blank screen before the beginning of the task. We also set a 3-minute time limit for participants to type each of their responses, with a beep sound played 1 minute before the time expired.

Table 1: Materials for the metaphor production task.

Action	Emotion
Revealing a secret is	Being happy is
Telling a lie is	Being ashamed is
Hurting someone is	Being sad is
Betraying a friend is	Being afraid is
Making a mistake is	Being angry is
Sleeping too much is	Being jealous is
Bothering friends is	Being surprised is
Spending all one’s savings is	Being interested is
Doing something useless is	Being envious is

**Music Stimuli** We selected two classical music pieces that differed in arousal and affective valence. The happy piece, *The Four Seasons*, Op. 8, No. 1, RV 269, Spring–Mvt 1. Allegro by Antonio Vivaldi (-27.05 dB), is characterized by high arousal and positive affect. The sad piece, *Adagio for Strings*, Op. 11 by Samuel Barber (-37.23 dB), is associated with low arousal and negative affect. Both pieces have been validated in prior research as reliably inducing the intended emotional states. Jefferies et al. (2008) observed that participants experienced significant mood changes, reporting increases in both affect and arousal while listening to the happy piece, and the opposite pattern while listening to the sad piece.

### Procedure

Participants completed both creativity tests and the metaphor production task in a single session. The tasks were administered in the following order: the Remote Associates Test (8 minutes), the Divergent Association Task (3 minutes), and the metaphor production task (20 minutes). The metaphor production task was conducted in one of three music types—silence, happy, or sad music—determined by their assigned version of the study.

After data collection, participant responses from the metaphor production task were evaluated through a categorization task and a rating task.

<sup>1</sup>Groups did not differ in age,  $F(2, 87) = 2.04, p = .14$ .

**Categorization Task** Three judges, blind to condition, categorized participants’ responses as either literal or figurative, assigning them to one of eight sub-categories. Literal responses included paraphrasing (PAR; rewording the original meaning), cause or consequence (CC; describing a cause or result), and example (EX; providing a concrete instance). Figurative responses included transparent metaphor (TM; easily interpretable metaphor), opaque metaphor (OM; complex or less accessible metaphor), simile (SI; explicit comparison using “like”), idiom (ID; fixed figurative expression), and synecdoche or metonymy (SM; part-whole or associative substitution). Judges received training in two live sessions that combined theoretical instruction with practical exercises.

**Rating Task** Two raters, independent from those in the categorization task, evaluated participants’ responses along four dimensions: comprehensibility, appropriateness, novelty, and valence. Each dimension was rated on a 7-point Likert scale. Comprehensibility referred to how easily the meaning of the response could be understood. Appropriateness referred to how well the response aligned with the prompt. Novelty referred to the originality and creativity of the response. Valence referred to the emotional tone of the response, ranging from negative to positive. Raters received detailed instructions and example responses for each criterion.

### Statistical Analysis

Analyses were conducted in R version 4.4.0 (R Core Team, 2024). Data processing and visualization were performed using the `tidyverse` package (Wickham et al., 2019), and Bayesian models with the `brms` package (Bürkner, 2017).<sup>2</sup>

We fitted a Bayesian logistic regression model to predict response category, using the Bernoulli family with a binary outcome variable (0 for literal, 1 for figurative). For model M1, weakly informative priors,  $N(0, 0.5)$ , were specified for all predictors. Based on Ritter and Ferguson (2017), which reported a small but significant effect between the silence and happy-music groups, we also fitted models M2, M3, and M4 with empirical directional priors for the happy-music group, assigning  $N(0.5, 0.1)$ ,  $N(0.5, 0.25)$ , and  $N(0.5, 0.5)$ , respectively to reflect prior evidence of a positive effect.

Then, we fitted four separate Bayesian ordinal regression models, using the cumulative family to model ratings (ranging from 1 to 7) of comprehensibility, appropriateness, novelty, and emotional valence. In all models, weakly informative priors,  $N(0, 0.5)$ , were used for all predictors.

For all analyses, music group and prompt type (emotion or action) were the predictors of interest, with RAT and DAT scores, trial order, and age as covariates. Random intercepts were included for items and participants. All models converged well ( $\hat{R} = 1.0$ ), and posterior predictive checks confirmed that models accurately predicted the data distribution.

<sup>2</sup>Data for this project, including processing scripts and analysis results, are available at [https://osf.io/5nfym/?view\\_only=b70fb6cbda45099143f15542b4c96a](https://osf.io/5nfym/?view_only=b70fb6cbda45099143f15542b4c96a).

## Results

This section presents descriptive statistics for the creativity measures, along with regression models for category (literal or figurative) and ratings (ranging from 1 to 7) of comprehensibility, appropriateness, novelty, and emotional valence.

### Descriptive statistics

**Creativity Measures** Creativity was assessed through two tasks, convergent thinking was measured using the Remote Associates Test (RAT) and divergent thinking was assessed through the Divergent Association Task (DAT). Table 2 presents RAT and DAT scores per group<sup>3</sup>. RAT scores reflect the number of correct responses out of 30 items (Cronbach’s  $\alpha = .92$ ). DAT measures the semantic distance between words, yielding a single creativity score per participant, with a theoretical range from 0 to 200, though scores typically fall between 65 and 90<sup>4</sup>. Following Olson et al. (2021), scores below 50 indicate poor performance, 75 to 80 reflect average performance, and scores above 95 indicate high performance.

We computed the Pearson correlation between RAT and DAT scores using the `Hmisc` package in R (Harrell Jr, 2024). The correlation result ( $R^2 = 0.11$ ,  $p = .28$ ) revealed a weak and non-significant relationship, suggesting that the two measures capture distinct components of creativity.

Table 2: Performance on the creativity tests across groups.

Group	RAT			DAT		
	M	SD	Range	M	SD	Range
Silence	10	7	0 – 24	77	5	60 – 85
Happy	11	8	0 – 26	78	5	68 – 86
Sad	12	8	0 – 27	78	4	69 – 85

**Response Category** First, judges categorized a shared set of 171 responses (10.5% of the data). Inter-rater reliability was measured using Krippendorff’s  $\alpha$  for nominal data, calculated with the `irr` package in R (Gamer et al., 2019). The resulting value was 0.70, indicating substantial agreement between the judges. Following this, judges were assigned unique sets of responses in two sessions. In the first session, each judge rated 300 sentences. In the second session, they rated an additional 180 sentences. Nine entries were excluded due to missing data.

Out of the total 1611 assigned responses, 85 (5.3% of the data) could not be categorized due to missing information and were marked as missing data. Additionally, 496 responses (32.3% of the data) were categorized as non-classified (NC). These responses included incomplete sentences (e.g., *being sad is part of*), sentences that did not adhere to the instructions (e.g., *being jealous is pointless in most cases*), and

<sup>3</sup>Groups did not differ in RAT,  $F(2, 87) = 0.82$ ,  $p = .44$ , or DAT scores,  $F(2, 87) = 0.66$ ,  $p = .52$ .

<sup>4</sup>DAT scores were calculated using the online scoring tool. For more information, visit <https://www.datcreativity.com>.

sentences that were nonsensical or did not fit any category (e.g., *hurting someone is never the answer*). NC responses were excluded from further analyses. Moreover, only two entries were categorized as synecdoche or metonymy (SM), and were excluded due to their low occurrence (0.1% of the data). Table 3 provides example responses from different participants to the same prompt, assigned to each category.

Table 3: Example responses for the prompt “Being happy is”.

Category	Prompt	Response
PAR	Being happy is	... <i>a state of joy</i>
CC	Being happy is	... <i>showing love</i>
EX	Being happy is	... <i>a walk on the beach</i>
TM	Being happy is	... <i>sunshine and rainbows</i>
OM	Being happy is	... <i>soaring on a cloud</i>
SI	Being happy is	... <i>like butterflies</i>
ID	Being happy is	... <i>sitting on cloud nine</i>

Overall, more entries were categorized as literal (546 responses, 52.7%) compared to figurative (491 responses, 47.3%). However, when comparing literal and figurative responses across music groups, the silence and sad-music groups exhibited a similar pattern, while the happy-music group showed an inverted pattern (see Figure 1) with more entries categorized as figurative (204 responses, 53.7%) compared to literal (176 responses, 46.3%).



Figure 1: Total response count across music groups.

**Response Ratings** First, raters evaluated a shared set of 110 responses (10.6% of the data). Inter-rater reliability for each dimension was assessed using the intraclass correlation coefficient (ICC), calculated with the `irr` package in R (Gamer et al., 2019) based on single-rating, consistency, and two-way mixed-effects models. The ICC values for comprehensibility (ICC = 0.64, 95% CI = [0.51, 0.74]), appropriateness (ICC = 0.64, 95% CI = [0.51, 0.73]), and novelty (ICC = 0.60, 95% CI = [0.46, 0.70]) indicated substantial agreement, and the value for valence (ICC = 0.82, 95% CI = [0.75, 0.88]) indicated strong agreement. Each rater was then assigned approximately 500 responses, representing half of the total set.

Table 4 presents the descriptive statistics and correlations for the ratings across all four dimensions for transparent and opaque metaphors. Overall, comprehensibility and appropriateness were strongly positively correlated. Novelty showed moderate negative correlations with both comprehensibility and appropriateness, supporting the notion that less coherent or suitable expressions are perceived as more original and creative. In contrast, valence showed no significant correlations with the other dimensions. These findings align with the predictions of Levorato and Cacciari (2002), but diverge from their empirical results, which reported positive correlations between comprehensibility, appropriateness, and novelty.

### Regression Models of Category

This analysis aimed to examine the probability of producing figurative sentences across the silence, happy-music, and sad-music groups. We predicted that participants listening to happy music would produce more figurative responses compared to those in silence. Additionally, we reported whether prompt type (emotion or action), RAT and DAT scores influenced the probability of producing figurative sentences.

The results of the regression models using different priors for the happy-music group are summarized in Table 5. In line with our prediction, all models showed a positive effect of happy music, indicating higher odds of producing figurative responses. In models with wider priors (M1 and M4), the credible intervals slightly overlapped with zero but the probabilities of a positive effect were large (M1:  $p(\hat{\beta} > 0) = 0.95$ ; M4:  $p(\hat{\beta} > 0) = 0.96$ ), and Bayes Factors (BF)<sup>5</sup> were greater than 3, indicating moderate evidence. Models with directional narrow (M2; Figure 2) or moderate priors (M3) excluded zero from the credible intervals and had Bayes Factors greater than 3, providing moderate evidence for the positive effect of happy music on figurative language production.

Across all models, the effect of prompt type was  $\hat{\beta} = 0.00$  (95% CRI = [-0.29, 0.30]), indicating that emotion and action prompts were equally effective in eliciting figurative sentences. Regarding measures of creativity, the effect of RAT scores was  $\hat{\beta} = 0.07$  (95% CRI = [0.04, 0.11]), suggesting that participants with higher convergent creativity had higher likelihood of producing figurative sentences. In contrast, the effect of DAT scores was  $\hat{\beta} = 0.00$  (95% CRI = [-0.05, 0.05]), suggesting no effect of divergent creativity.

### Regression Models of Ratings

This analysis aimed to examine whether metaphorical responses differed in comprehensibility, appropriateness, novelty, and valence across groups. We predicted that metaphors from the happy-music and sad-music groups would be rated as more novel than those from the silence group, and that

<sup>5</sup>Bayes Factors (BF) quantify the strength of evidence for one model over another. A BF below 3 indicates anecdotal evidence, between 3 and 10 suggests moderate evidence, and above 10 indicates strong evidence (Jeffreys, 1961). We compared a full model (with music effects) to a reduced model (without music effects) across varying standard deviations for the happy-music group.

Table 4: Statistics and correlations for ratings of comprehensibility (C), appropriateness (A), novelty (N), and valence (V).

	Transparent Metaphors						Opaque Metaphors					
	Statistics		Correlations				Statistics		Correlations			
	M	SD	C	A	N	V	M	SD	C	A	N	V
Comprehensibility	5.73	1.73	-				4.63	2.35	-			
Appropriateness	5.58	1.86	.89*	-			4.40	2.35	.94*	-		
Novelty	4.10	2.12	-.57*	-.57*	-		5.50	1.80	-.53*	-.51*	-	
Valence	3.11	1.69	.08	.03	-.05	-	3.01	1.33	-.14	-.15	-.08	-

Table 5: Regression coefficients for literal and figurative categories across different priors for the happy-music group.

	M1		M2		M3		M4	
	$N(0, 0.5)$		$N(0.5, 0.1)$		$N(0.5, 0.25)$		$N(0.5, 0.5)$	
	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI
Silence	-0.37	[-0.77, 0.02]	-0.42	[-0.75, -0.10]	-0.43	[-0.79, -0.07]	-0.39	[-0.78, -0.00]
Happy	0.41	[-0.07, 0.89]	0.50	[0.32, 0.69]	0.52	[0.15, 0.88]	0.43	[-0.05, 0.91]
Sad	-0.21	[-0.71, 0.29]	-0.17	[-0.62, 0.29]	-0.16	[-0.64, 0.32]	-0.20	[-0.70, 0.29]
BF	3.34		9.96		7.94		3.89	

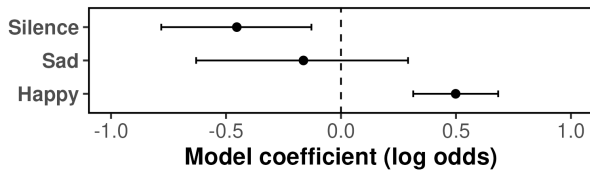


Figure 2: Posterior distributions for model M2.

metaphors for emotions would be rated as more comprehensible and appropriate but less novel than those for actions. Additionally, we reported whether prompt type, RAT, or DAT scores influenced these ratings. We also verified the category judgment results by assessing whether transparent metaphors were rated higher for comprehensibility and appropriateness but lower for novelty compared to opaque metaphors.

The results of the regression models are summarized in Table 6. In line with our first prediction, metaphors from the happy and sad-music groups were generally rated as more novel compared to those from the silence group. Although the 95% credible intervals included zero, there were high probabilities of positive effects for both the happy-music group,  $p(\hat{\beta} > 0) = 0.87$ , and the sad-music group,  $p(\hat{\beta} > 0) = 0.75$ .

Moreover, metaphors from the sad-music group were generally rated as more negative in emotional valence (1 for negative to 7 for positive) than those from the silence group. Although the 95% credible interval slightly included zero, the probability of a negative effect was large,  $p(\hat{\beta} < 0) = 0.94$ .

In line with our second prediction, metaphors expressing emotions were rated as more comprehensible and appropriate but less novel than those used for actions (see Table 6).

Additionally, transparent metaphors were rated higher in

comprehensibility ( $\hat{\beta} = 0.71$ , 95% CRI = [0.36, 1.07]) and appropriateness ( $\hat{\beta} = 0.80$ , 95% CRI = [0.45, 1.16]) but lower in novelty ( $\hat{\beta} = -1.06$ , 95% CRI = [-1.41, -0.69]) than opaque metaphors. This confirmed that judges distinguished between transparent and opaque metaphors in the categorization task.

Lastly, RAT and DAT scores showed no significant effects across any of the models, with estimates near zero and wide credible intervals, indicating no predictive power.

## Discussion

We investigated the extent to which metaphor production is influenced by different music-induced mood states. First, we examined the probability of producing figurative expressions during a metaphor production task performed in silence, with happy music, or with sad music in the background. Then, we assessed the quality of the responses by evaluating their comprehensibility, appropriateness, novelty, and valence.

Building on research indicating that listening to happy music enhances divergent thinking (Ritter & Ferguson, 2017), we hypothesized that it would also facilitate the generation of metaphors. Our results support this hypothesis, showing that only participants who performed the task with happy music in the background were more likely to produce figurative responses than literal ones. This finding suggests that metaphor production relies on a component of creativity that is enhanced by happy music, but not by sad music.

Our results are consistent with Ashby et al. (1999)'s theory of positive affect, which proposes that a slight increase in positive feelings facilitates the release of dopamine in the brain, enhancing cognitive flexibility and thereby improving performance on creative tasks. More recently, Nijstad et al. (2010) developed the dual pathway model of creativity, which posits that creativity arises from two different processes, cog-

Table 6: Regression coefficients for ratings of comprehensibility, appropriateness, novelty, and valence.

	Comprehensibility		Appropriateness		Novelty		Valence	
	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI	$\hat{\beta}$	95% CRI
Happy	-0.38	[-0.85, 0.10]	-0.39	[-0.87, 0.10]	0.26	[-0.21, 0.71]	-0.14	[-0.53, 0.24]
Sad	0.04	[-0.47, 0.55]	-0.01	[-0.54, 0.52]	0.17	[-0.32, 0.65]	-0.34	[-0.77, 0.09]
Emotion	0.49	[0.12, 0.86]	0.49	[0.12, 0.86]	-0.67	[-1.03, -0.31]	0.18	[-0.37, 0.75]

nitive flexibility and cognitive persistence. The flexibility pathway refers to the ability to form new associations and adopt novel strategies for solving open-ended problems. In contrast, the persistence pathway involves the sustained exploration of multiple options within a limited conceptual domain. Both of these pathways can be influenced by situational factors, such as exposure to unusual events (Ritter et al., 2012) or listening to happy music (Ritter & Ferguson, 2017). In the metaphor production task, it is possible that the positive mood induced by the high arousal levels of the happy music increased cognitive flexibility, enabling participants to connect the prompted actions and emotions to experiences from more distant or unrelated domains (e.g., *being surprised is a sudden spark of electricity*) rather than resorting to more literal responses grounded in similar domains (e.g., *being surprised is being excited and positive*).

We further hypothesized that metaphors produced with background music would be perceived as more novel (i.e., more original and more creative) compared to those completed in silence. In line with this prediction, metaphors produced with background music—particularly in the happy-music group—were generally perceived as more novel, compared to those completed in silence. A recent study by Winter and Strik-Lievers (2023) showed that the creativity of metaphors can be explained by the semantic distance between their components; metaphors with greater semantic distance (e.g., *loud color*) tend to be perceived as more creative than those with lower semantic distance (e.g., *bright color*). This finding further supports the idea that increased cognitive flexibility induced by music enabled participants to produce metaphors involving more semantically distant domains, which were, in turn, rated as more creative.

Moreover, participants with higher convergent creativity were more likely to generate figurative responses, whereas no differences were observed with respect to divergent creativity. These findings do not necessarily contradict those of Menashe et al. (2020), who suggested that divergent thinking contributed to the generation of novel responses. Their task specifically prompted the generation of similes (e.g., X is like Y), which are explicit comparisons, whereas ours allowed for the generation of metaphors (e.g., X is Y), which are categorical statements. Given that metaphors and similes yield different semantic representations (Roncero et al., 2021), with similes more frequently used for novel comparisons (Roncero et al., 2006), these differences in task design may account for the contrasting results. It is also important to note

that Menashe et al. (2020) employed the Tel Aviv Creativity Test (TACT; Milgram & Milgram, 1976), while we employed the Divergent Association Task (DAT) with no studies directly comparing the two. In addition, the DAT provides a single overall score per participant, preventing the calculation of internal consistency reliability, whereas we obtained high reliability in the RAT. Another possible explanation for the observed pattern is task order. Participants completed the convergent task first, followed by the divergent task, which may have influenced their performance by constraining their thinking during the latter task. Although the effects of these measures were not the focus of our study, these differences may explain why convergent creativity, but not divergent creativity, influenced the likelihood of generating figurative responses during the production task.

This study also examined the use of metaphors to express emotions versus actions. Based on previous findings (Levorato & Cacciari, 2002), we hypothesized that emotion prompts would elicit more metaphorical and suitable responses than action prompts. Our results partially supported this prediction. While both prompts were equally likely to elicit figurative responses, metaphors for emotions were perceived as more comprehensible and appropriate but less novel than metaphors for actions. This pattern suggests that metaphors effectively capture the richness and complexity of emotional internal states, whereas actions, being more concrete and externally observable, are often better expressed through literal descriptions (Fainsilber & Ortony, 1987).

This study extends findings on the effects of happy music on divergent thinking tests to an everyday task—metaphor production. While previous research has examined how happy music influences performance on standardized assessments of creativity, our study shows that happy music facilitates figurative language production in a more natural, less structured creative task. This highlights the impact of extrinsic factors on creativity and suggests potential applications for enhancing creative thinking in real-world contexts.

Future research should explore alternative methods of inducing positive mood—such as visual stimuli, social interactions, or autobiographical recall—and examine how mood fluctuates at different stages of the task. Additionally, the relationship between creativity and metaphor production could be further investigated using more controlled tasks, such as generating metaphorical vehicles for a broader range of topics, to better understand the cognitive processes underlying the generation of creative metaphors.

## Acknowledgments

This research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 948878).

We thank Yildiz Aslan and Baoyu Xiong for their contributions to the study setup and their role as judges in the categorization task. We are also grateful to Niranjana Satyanarayana for programming the experiment and to Sophia Wiedmann and Eva Gavaller for serving as raters in the rating task. We also thank Chris Genovesi for his helpful comments on an earlier version of this article.

## References

- Adaman, J. E., & Blaney, P. H. (1995). The effects of musical mood induction on creativity. *The Journal of Creative Behavior, 29*(2), 95–108.
- Al-Azary, H., & Buchanan, L. (2017). Novel metaphor comprehension: Semantic neighbourhood density interacts with concreteness. *Memory & Cognition, 45*, 296–307.
- Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological review, 106*(3), 529.
- Beatty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory & cognition, 41*, 255–267.
- Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers, 35*, 634–639.
- Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software, 80*(1), 1–28.
- Callaghan, K. T., & Growney, C. M. (2013). The impact of music and mood on creative thinking. *Psi Chi Journal of Psychological Research, 18*(4).
- 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.
- Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol, 2*(4), 239–250.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *Irr: Various coefficients of interrater reliability and agreement*.
- Guilford, J. (1950). Creativity. *American Psychologist, 5*(9), 444–454.
- Harrell Jr, F. E. (2024). *Hmisc: Harrell miscellaneous* [R package version 5.1-3].
- Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of personality and social psychology, 52*(6), 1122.
- Jefferies, L. N., Smilek, D., Eich, E., & Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychological Science, 19*(3), 290–295.
- Jeffreys, H. (1961). *Theory of probability* (3rd). Oxford University Press.
- Levorato, M. C., & Cacciari, C. (2002). The creation of new figurative expressions: Psycholinguistic evidence in Italian children, adolescents and adults. *Journal of Child Language, 29*(1), 127–150.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232.
- Menashe, S., Leshem, R., Heruti, V., Kasirer, A., Yair, T., & Mashal, N. (2020). Elucidating the role of selective attention, divergent thinking, language abilities, and executive functions in metaphor generation. *Neuropsychologia, 142*.
- Milgram, R. M., & Milgram, N. A. (1976). *Tel aviv creativity test (TACT)*. Tel Aviv University, School of Education.
- Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European review of social psychology, 21*(1), 34–77.
- Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences, 118*(25), e2022340118.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Ritter, S. M., Damian, R. I., Simonton, D. K., van Baaren, R. B., Strick, M., Derks, J., & Dijksterhuis, A. (2012). Diversifying experiences enhance cognitive flexibility. *Journal of experimental social psychology, 48*(4), 961–964.
- Ritter, S. M., & Ferguson, S. (2017). Happy creativity: Listening to happy music facilitates divergent thinking. *PloS one, 12*(9), e0182210.
- Roncero, C., De Almeida, R. G., Pissani, L., & Patalas, I. (2021). A metaphor is not like a simile: Reading-time evidence for distinct interpretations for negated tropes. *Metaphor and symbol, 36*(2), 85–98.
- Roncero, C., Kennedy, J. M., & Smyth, R. (2006). Similes on the internet have explanations. *Psychonomic Bulletin & Review, 13*, 74–77.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., & François, R. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.
- Winter, B., & Strik-Lievers, F. (2023). Semantic distance predicts metaphoricity and creativity judgments in synesthetic metaphors. *Metaphor and the Social World, 13*(1), 59–80.