Creative fixation is no laughing matter: The effects of funny and unfunny examples on humor production

By: Hyelim Shin, Katherine N. Cotter, Alexander P. Christensen, and Paul J. Silva

This is the peer reviewed version of the following article:


which has been published in final form at https://doi.org/10.1002/jocb.383. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

***© Creative Education Foundation. Reprinted with permission. No further reproduction is authorized without written permission from Wiley. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. ***

Abstract:

How do people come up with humorous ideas? In creative cognition research, exposure to good examples sometimes causes fixation (people get “stuck” on the examples) but other times sparks inspiration (people's responses are more creative). The present research examined the effects of funny and unfunny examples on joke production. A sample of 175 adults read scenarios that they completed with funny responses. All participants were instructed to be funny, but before responding they read (a) funny responses as examples of good responses to emulate, (b) unfunny responses as examples of poor responses to avoid, or (c) no examples. The participants’ own responses were rated for funniness and for similarity to the example responses, and response times were recorded. Reading either funny or unfunny examples, compared to no examples, caused people to come up with funnier jokes. Similarity to the examples was low in all conditions, so fixation was relatively modest, but people who saw unfunny examples spent more time coming up with their responses. Taken together, the findings support the growing literature showing that examples are often inspiring rather than constraining, and they imply that good and bad examples spark creative thought via different paths.

Keywords: humor | creativity | fixation | inspiration | joke production

Article:

It is almost funny how little is known about the cognitive psychology of humor. Verbal humor—funny ideas expressed via language—is an interesting example of creativity. Humor is ubiquitous in everyday life: the quips, puns, jokes, and ripostes of daily speech are excellent examples of mini-c and little-c creativity (Kaufman & Beghetto, 2009) and illustrate the essentially creative quality of informal social interaction (Cotter, Christensen & Silvia, 2018; Tanggaard, 2015). The
cognitive processes and mechanisms involved in generating jokes and witticisms are probably much like those involved in common creativity paradigms. Coming up with funny ideas involves reconfiguring one's knowledge, thinking with analogy and metaphor, making implicit knowledge explicit, and digging deep into the different shades and senses of words and concepts (Earleywine, 2010; Goatly, 2012).

To assess humor production, researchers use open-ended tasks that resemble common creativity tasks. The participants are given a prompt, such as a cartoon or quirky scenario, and then generate something funny, such as the ending to a set-up, a caption for a cartoon, or a definition for an ambiguous concept (Nusbaum & Silvia, 2017). The responses are then rated for funniness by a group of judges (Nusbaum, 2015). Based on the growing body of work, humor production tasks resemble popular creativity tasks in their network of relationships. Like divergent thinking and metaphor production (Silvia, 2015), for example, intelligence predicts the ability to come up with funny ideas. Both fluid intelligence and crystalized intelligence are correlated with humor production (Christensen, Silvia, Nusbaum & Beaty, 2018; Greengross & Miller, 2011; Howrigan & MacDonald, 2008), and professional comedians have significantly higher vocabulary scores than university students (Greengross, Martin & Miller, 2012). Openness to experience—an important trait in virtually all creativity tasks (Oleynick et al., 2017)—predicts higher scores on humor production tasks (Nusbaum & Silvia, 2017), consistent with the imaginative, unconventional, and intellectual qualities of people high in openness (Christensen, Cotter & Silvia, 2018).

Nevertheless, the cognitive mechanisms involved in generating funny ideas are much less understood than the mental nuts-and-bolts of popular paradigms, such as generating unusual uses, remote associates, or metaphors. To extend the cognitive psychology of humor, we examined the role of fixation—a salient process in many other creative contexts—when people try to be funny. Fixation research examines how prior knowledge can interfere with creative thinking for both convergent and divergent tasks. Research on problem-solving and insight, for example, has shown that knowledge of an object's common function, surface properties, or salient features can impede finding a correct solution (Duncker, 1945; McCaffrey, 2012). In idea generation tasks, which lack correct answers, fixation research has explored how example responses affect people's own ideas (Smith, Ward & Finke, 1995). In classic early studies, participants worked on open-ended design tasks, such as designing toys and drawing aliens. Participants who saw examples typically included features of the examples in their own designs, even when told to avoid copying their features (Smith, Ward & Schumacher, 1993). These studies suggested that prior examples can constrain creativity by providing salient concepts and features that intrude into subsequent thought (Smith & Blankenship, 1991).

Like most effects in psychological research, the fixation effect has proved to be complex. For one, showing people prior examples does not necessarily harm the creativity of their ideas. A recent meta-analysis concluded that people typically do include features of the examples in their own responses, but their responses are more novel as a result (Sio, Kotovsky & Cagan, 2015). Experiments suggest that examples can inspire more creative responses by implying analogies, ideation strategies, or good starting points (Agogué et al., 2014; George, Wiley, Koppel & Storm, 2018; Marsh, Landau & Hicks, 1996).

Bad examples—ideas to avoid—are an intriguing wrinkle in fixation studies. A small group of studies has given participants examples of mediocre ideas, such as responses that are common
(Fink et al., 2012; Yagolkovskiy & Kharkhurin, 2016) or conceptually close (Agogué et al., 2014). In their meta-analysis, Sio et al. (2015) excluded studies of what they termed “negative examples” from their formal analysis because they noted that relatively few studies that presented negative examples measured eventual design quality or novelty. Some studies outside of design, however, have found that bad examples (e.g., common divergent thinking responses; Fink et al., 2012) can spark better ideas from participants. Instead of including features of the bland or common responses, for example, people seem to inductively draw something useful from them, such as features or starting points to avoid.

In the present research, we examined the role of examples in humor. Our study sought to extend fixation research to humor production, a novel domain of creativity. Fixation is intriguing for humor not only because it has yet to be studied, but because the solution space for humor tasks is vast. In humor tasks, a response need only strike the respondent as funny. Because the requirements for feasibility, appropriateness, or aptness are so loose in humor—funny things are often bizarre and inane—the range of potential responses is enormous. Examples might thus exert a large influence in humor by providing people a place to start. As a result, it seems likely that the effects of examples, whether positive or negative, would be exaggerated. We particularly sought to focus on the distinction between positive and negative examples—funny and unfunny responses taken from our past studies, in this case—to see if they have different effects on people's attempts to be funny. We did this by presenting participants with either funny examples, unfunny examples, or no examples and then examining their own responses to the same humor production tasks. These responses were also evaluated for evidence of fixation, assessed via the similarity of the responses to the examples in key words, concepts, and themes.

**Method**

**Participants and Procedure**

Participants were 175 students attending the University of North Carolina at Greensboro (UNCG). The sample was young (\(M = 18.75\) years old, \(SD = 0.99\)), predominately female (77.2% female), and diverse (38% African American, 8% Hispanic/Latino(a), 49% European American). Participants were randomly assigned to 1 of 3 between-group conditions: no examples, unfunny examples, or funny examples. Participants completed the study using MediaLab software on desktop computers in small groups. All participants provided informed consent, and the study was approved by our institution's IRB.

**Humor Task and Example Manipulation**

To assess humor production, we used a joke stem task (Christensen, Silvia et al., 2018; Nusbaum, Silvia & Beaty, 2017). In this task, participants are given two funny scenarios as set-ups and then are asked to write a funny ending for each prompt. Much like “be creative” instructions for divergent thinking tasks (Nusbaum, Silvia & Beaty, 2014), the joke stems task used “be funny” instructions that emphasized coming up with humorous responses:

For this next task, you will be presented with the first part of a joke. Your job is to finish the joke by writing something funny. It can be weird, silly, strange, dirty-minded, ironic, or whatever else, so long as it is funny.

Table 1 lists the two scenarios, which set people up with common contexts for college-student humor, and the full instructions. For example, the first scenario said, “Imagine that one of your
classes this semester is incredibly boring, and you're trying to convey just how incomprehensibly boring this class is to one of your friends. So, you say, ‘Seriously, that class is so boring …’” Participants were asked to complete the sentence with their own humorous response. They could take as long as they wanted for each prompt, but they provided only one response. The software recorded their response time, measured in seconds from the item's onset to when participants submitted their response.

<table>
<thead>
<tr>
<th>Funny examples</th>
<th>Unfunny examples</th>
<th>No examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boring Class Scenario</strong></td>
<td>Imagine that one of your classes this semester is incredibly boring, and you're trying to convey just how incomprehensibly boring this class is to one of your friends. So, you say, “Seriously, that class is so boring, …” You should complete this joke by writing something funny. Try to write fun and silly ideas, like “… I could bring boring to the class and boring would leave,” “… it could put the History Channel to sleep,” and “… it was like being at a funeral when I didn't know who died.”</td>
<td>Imagine that one of your classes this semester is incredibly boring, and you're trying to convey just how incomprehensibly boring this class is to one of your friends. So, you say, “Seriously, that class is so boring, …” You should complete this joke by writing something funny. Try to avoid boring and obvious ideas, like “… it put the teacher to sleep,” “… it was like watching paint dry,” or “… it was like watching grass grow.”</td>
</tr>
<tr>
<td><strong>Gross Food Scenario</strong></td>
<td>Imagine that your friend invites you over and cooks dinner—and the food is totally horrible and disgusting. Later, when describing it to someone else, you say, “Wow, that food was so bad …” You should complete this joke by writing something funny. Aim for fun and silly ideas like “… it felt like World War III was going on inside my stomach,” “… my taste buds fell out of my mouth and started weeping,” and “… my vagina hurt.”</td>
<td>Imagine that your friend invites you over and cooks dinner—and the food is totally horrible and disgusting. Later, when describing it to someone else, you say, “Wow, that food was so bad …” You should complete this joke by writing something funny. Try to avoid boring and obvious ideas, like “… the dog wouldn't eat it,” “… it tasted like vomit,” or “… I could have died.”</td>
</tr>
</tbody>
</table>

The software randomly assigned the participants to receive funny examples, unfunny examples, or no examples. The condition applied to both tasks (i.e., people in the funny-examples condition received funny examples for both joke stems). Table 1 lists the examples that participants read. For realism, we took the funny and unfunny examples from responses collected from participants in previous studies (Christensen, Silvia et al., 2018; Nusbaum et al., 2017). The funny examples had some of the highest humor ratings in their original study; the unfunny examples were common responses that always received the lowest humor ratings. The funny and unfunny examples were explicitly marked as being good or bad instances of the kinds of responses to generate. For the funny examples, for example, people read:

You should complete this joke by writing something funny. Try to write fun and silly ideas, like “… I could bring boring to the class and boring would leave,” “… it could put the History Channel to sleep,” and “… it was like being at a funeral when I didn't know who died.”

In the unfunny condition, in contrast, people read:
You should complete this joke by writing something funny. Try to avoid boring and obvious ideas, like “… it put the teacher to sleep,” “… it was like watching paint dry,” or “… it was like watching grass grow.”

Scoring the Responses

The participants’ responses were subjectively coded for two qualities. First, as in past research, we scored how funny the responses were. Three raters scored each response using a 5-point scale (1 = not at all funny, 5 = very funny). The raters were unaware of a participant's condition, the other raters’ scores, and the participant's other data. The humor scores were averaged across tasks and raters to yield a single funniness score per person (α = .71).

Second, to assess fixation, the raters scored how similar the participant's response was to the example responses that they viewed. This rating was done only for the funny- and unfunny-examples conditions. (Ratings for responses in the no-examples condition would reflect whether their responses, by chance, happened to resemble examples other participants saw.) For each condition, raters were given the examples of funny and unfunny responses and instructed to rate each participant's response using a 3-point scale:

- 0 = No obvious overlap. The response does not copy key words or ideas.
- 1 = Some overlap. The response might copy a key word or idea but expands or develops it.
- 2 = A copy, more or less. The response is mostly the same idea and words.

Higher scores indicate greater overlap with the examples that were presented in that condition (i.e., how much responses in the good-examples condition resembled the funny examples and responses in the unfunny-examples condition resembled the unfunny examples). The raters were unaware of the other raters’ scores and the participant's other data. The similarity scores were averaged across raters and tasks to yield a single similarity score per participant (α = .70).

Results

Analysis Approach

The humor ratings were analyzed using polynomial contrasts, which estimated linear (−1, 0, 1) and quadratic (1, −2, 1) effects across the three groups. The linear order of the conditions was unfunny examples (−1), no examples (0), and funny examples (1). The linear contrast thus evaluates predictions of a fixation effect: good examples spark funnier responses, and bad examples spark worse responses. The quadratic contrast evaluates whether any examples, good or bad, affect the funniness of people's responses relative to no examples. This contrast thus reflects the possibility that bad examples can also improve funniness, consistent with some past fixation work. Differences between conditions were followed up using 95% confidence intervals around their means.

The analyses were conducted in Mplus 8.1 using maximum likelihood estimation with robust standard errors. All regression coefficients are standardized and can be interpreted within the r effect size metric (.10, .30, .50 as small, medium, and large, respectively; Cumming, 2012). Table 2 shows the descriptive statistics and confidence intervals for the outcomes.
Table 2. Descriptive Statistics for the Unfunny-Examples, No-Examples, and Funny-Examples Conditions

<table>
<thead>
<tr>
<th></th>
<th>Unfunny examples</th>
<th>No examples</th>
<th>Funny examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>95% CI</td>
</tr>
<tr>
<td>Funniness rating</td>
<td>1.85</td>
<td>0.53</td>
<td>1.71, 1.99</td>
</tr>
<tr>
<td>Closeness to example</td>
<td>0.19</td>
<td>0.24</td>
<td>0.13, 0.25</td>
</tr>
<tr>
<td>Response time (raw)</td>
<td>57.63</td>
<td>32.68</td>
<td>48.88, 66.39</td>
</tr>
<tr>
<td>Response time (log-transformed)</td>
<td>3.90</td>
<td>0.57</td>
<td>3.75, 4.05</td>
</tr>
</tbody>
</table>

Notes: n = 56 for unfunny examples group; n = 59 for no examples group; n = 60 for funny examples group.

Figure 1. Effects of the examples manipulation on (a) the funniness of the participants’ responses (top); (b) the similarity of the participants’ responses to the examples (middle); and time (in seconds, log-transformed) to generate a response (bottom). The bars represent 95% confidence intervals around the mean.

Humor Ratings

Did showing people funny or unfunny examples affect the funniness of their own responses? The regression model ($R^2 = 11.6\%$) for subjective humor ratings found a significant quadratic effect
(β = .34, p < .001) and no linear effect (β = .00, p = .969). Figure 1 (top panel) displays the pattern; Table 2 has the full descriptive statistics. Seeing any examples, good or bad, caused significantly funnier responses. According to the 95% confidence intervals, the funny- and unfunny-examples conditions did not differ from each other in humor ratings, but each condition was significantly higher than the no-examples condition.

Similarity to Examples and Response Time

Because both funny and unfunny examples caused funnier responses, a natural question is whether people's responses resembled the examples they saw, consistent with a fixation effect. (Recall that similarity ratings were scored for only the funny- and unfunny-examples condition, so the predictor has only two levels.) Ratings of similarity to the example responses were near the floor of zero in both the funny examples (M = 0.19, SE = .03) and unfunny-examples (M = 0.19, SE = .03) condition. A regression model (R² = 0.00%) for ratings of similarity to the examples found a near-zero effect (β = .00, p = .97); Figure 1 (middle panel) displays the pattern of results. The similarity of people's responses to the examples that they read was identical in the funny- and unfunny-examples condition (see Table 2).

Finally, we examined response times, which were log-transformed to make the scores more normally distributed. The polynomial regression model (R² = 3.9%) found a significant linear effect for response time (β = −.15, p = .029) and a marginal quadratic effect (β = .13, p = .096). Figure 1 (bottom panel) displays the pattern of results. The unfunny-examples condition was significantly slower than the other two (see Table 2), so reading bad examples caused people to take longer to create their own response. The no-examples and funny-examples conditions did not differ from each other.

Discussion

When do examples help humor? In the present experiment, we examined the effect of good and bad examples on people's own attempts to be funny. Past work, most of which has used design tasks, has found evidence for both fixation and inspiration (Sio et al., 2015). On the one hand, people do tend to include features of good examples in their own ideas, suggesting fixation. But on the other hand, seeing good examples tends to make their own ideas more novel and worthwhile than they otherwise would be (George et al., 2018), suggesting inspiration.

In the present experiment, the evidence showed an inspiring effect of examples on humor. When people were given good examples or bad examples, their own responses were significantly funnier compared to people in a no-examples control group. The effects of good and bad examples on humor were similarly strong, so a natural next question is how such different examples caused similar increases in funniness. Differences in fixation were probably not the reason. People in the funny- and unfunny-examples conditions showed identical levels of fixation, assessed as the rated resemblance of their response to any of the examples they read. Indeed, the absolute levels of rated similarity were very low, so the participants in general were relatively unlikely to fixate on the examples.

Past work has shown that examples can spark better responses for a few reasons, such as providing people with a good place to start or a fertile category that they can explore deeply (George et al., 2018; Sio et al., 2015). In the domain of humor, we suspect that examples additionally imply task norms and strategies that participants discern inductively and then apply
to the task. For example, good examples provide normative information about what kinds of responses are desirable and acceptable. If the good examples are bizarre, perverse, or elaborate—as some were in our study—participants probably infer that such responses are acceptable and will then delve into possible responses that they would have previously discarded as being too weird or dirty-minded. The tacit normative information conveyed by examples likely has a much larger effect in humor tasks than in other paradigms, such as design and divergent thinking tasks.

Likewise, examples can imply fruitful task strategies. Many studies, primarily on divergent thinking, have shown that giving participants good strategies usually improves the creative quality of their responses (e.g., Forthmann, Wilken, Doebler & Holling, 2018; Nusbaum & Silvia, 2011). Other studies have assessed how people generate and shift strategies over the course of a task (Gilhooly, Fioratou, Anthony & Wynn, 2007). Many of the strategies people use are unproductive, but others are more fertile. How people generate and select strategies is an intriguing and understudied question, but it seems likely that task features can tacitly suggest strategies. In this study, for example, the bad examples were among the most common and intuitive responses participants give. Presenting them as bad examples to avoid implies fruitful task strategies, such as setting aside the first thing that comes to mind or developing responses that are more elaborate than simple adjectival phrases (e.g., “it was totally nasty”). The differences in response times suggest a role for implied strategies. People who read the bad examples took significantly longer to generate their responses compared to the other two conditions. This is consistent with the possibility that bad examples caused people to go beyond their initial, first responses, which likely resembled one of the bad examples, and instead apply new strategies for generating a funny response. A major goal for future work would be to unpack the strategies that people apply to these open-ended humor tasks and how well they can be learned and used by people for whom humor comes less naturally.

References


(Eds.), *Cambridge handbook of creativity and personality research* (pp. 294–322). New York: Cambridge University Press. https://doi.org/10.1017/9781316228036


