

# Creativity in humor production: Quantity and quality in divergent thinking

PETER DERKS and DEDRECK HERVAS  
*College of William and Mary, Williamsburg, Virginia*

Humorous captions for still pictures from old movies were produced by 38 college students. Either 2 or 10 captions for each of three pictures were required. There was a less than 1% overlap among the captions, demonstrating that the task results in highly divergent answers. Average caption funniness improved with output order, indicating that in a task leading to divergent responses, quantity can lead to quality.

Weisberg (1986), in his critique of "myths" of creativity, presented evidence that sheer quantity of ideas does not always lead to original answers to problems. Productivity and brainstorming are not necessarily sources of creative solutions.

Nevertheless, as ideas are generated in unusual uses tasks, they tend to increase in uniqueness (Christensen, Guilford, & Wilson, 1957; Meadow & Parnes, 1959). Parnes (1961) also rated the "value" of unusual uses, but added the score to uniqueness for the final evaluation. Thus, even though quality increased with output, it was confounded by associative remoteness, which, almost by definition, should increase with increased output.

Consequently, the measure of the quantity-quality output relation requires either pure quality evaluation or a task that produces only unique responses. Since the judgment of quality of an unusual use must depend, to some extent, on its uniqueness, the second approach was attempted in the present study.

Several writers have proposed humor and humor production as an "act of creation" (Arietti, 1976; Koesler, 1964). Sense of humor and creativity even correlate somewhat (Brodzinsky & Rubien, 1976; Colell & Domino, 1980; Treadwell, 1970). Finally, when subjects were asked to write "funny" captions for still pictures from old movies, there was less than 1% overlap among captions (Derks, in press). Humor production is, therefore, a good candidate for a task that generates non-overlapping divergent responses.

Furthermore, humor production is a problem-solving task with a clear but difficult solution. The attempt at humor cannot be a direct association or completely unrelated to the situation. The best humor will arouse, surprise, resolve, and relate to the audience (McGhee, 1979; Ziv, 1984). Humor production is difficult, yet subjects from schizophrenics to college students are willing to attempt it (Derks, Leichtman, & Carroll, 1975; Rosin & Cerbus, 1984).

An earlier study that examined humor production did not find a general relation between output order of the captions and their quality (i.e., funniness) (Derks, in press). In that study, however, subjects were allowed to make as many or as few responses as they wished in a set time period. The variability in quantity and in latency of responses made it difficult to apply any parametric analysis. Nevertheless, "best" captions did tend to occur after "worst" by about two to one.

In the present study the humor producers were required to write a certain number of captions for the pictures. Therefore, "early" and "late" could be quantified more precisely. Performance on a large number of captions was also compared with performance on a small number. From this comparison, inferences could be made about how people monitored their idea production.

## METHOD

### Subjects

Thirty-eight introductory psychology students participated in humorous caption writing to fulfill a course requirement. They were tested in groups of 5 to 7.

Ten judges were also selected from the introductory class. They were chosen because their judgments of humor were consistent with a larger sample of students (Derks, Lewis, & White, 1981). These judges were paid \$10 for evaluating the 684 captions that resulted from humor production. The evaluation was done at their own paces on their own time.

### Materials

Three pictures were selected because they had elicited a relatively large number of different responses in previous research. They were Gibson Gowland chewing on the fingers of a screaming Zasu Pitts in a scene from "Greed," Lon Chaney in "The Phantom of the Opera" pointing gleefully to the left as Norma Philbin looks on with dismay, and Richard Dix making a rabbit shadow on the wall for a bemused and bedridden young woman in "A Man Must Live" (Franklin, 1959). A large digital clock permitted the subjects to record the time at which each caption occurred to them.

### Procedure

The pictures were projected one at a time in a random order. Nineteen subjects were instructed to write 10 captions, and the remaining subjects wrote 2 for each picture. These captions were to be as "funny as possible." The subjects also were told to record the time when the idea for the caption occurred to them. A picture was not removed until all subjects were finished.

---

Requests for reprints should be sent to Peter Derks, Department of Psychology, College of William and Mary, Williamsburg, VA 23185.

The judges were given a packet of the pictures and booklets of the relevant captions, which had been typed and randomized over subjects and order. They judged the captions on a 21-point scale, with 0 = *not at all funny* and 10 = *extremely funny*. Such a scale has been found to correlate well with other measures of humor (Derks et al., 1981).

## RESULTS

### Overlap

Of the 684 captions generated for the three pictures, 3 were verbatim repetitions. For example, the cliché "Not tonight, I have a headache" was given twice for "The Phantom of the Opera" picture. "Not tonight, I have a hangnail" was also given for the man chewing the lady's fingers, but was not counted as a repetition. The overlap rate was well below 1%.

### Quantity

Figure 1 presents humor rating (quality) as a function of order and rate of output (latency). When the subjects wrote 10 captions, the latencies got progressively longer [ $F(9,162) = 9.68, p < .0001$ ]. This effect was primarily due to the short average time, only 19 sec, required to think of the first caption. After that, the next shortest latency, from the second to the third caption, was 42 sec, and the longest latency, for the last caption, was 60 sec. The latencies for the three different pictures were significantly different [ $F(2,36) = 5.48, p < .1$ ] and interacted with output order [ $F(18,324) = 2.03, p < .01$ ]. Since these effects were of little importance to the present study, the latencies were combined for presentation in Figure 1. Of major importance is the consistent high rate of responding when a large number of captions was required.

The latency for providing only 2 captions was much slower than that for providing the first 2 of 10 [ $F(1,36) = 30.01, p < .0001$ ]. The subjects took as long to provide the second caption of 2 as they did to provide the third and fourth captions of 10. The subjects appeared to be more self-critical when they had fewer captions to write. Again latency slowed [ $F(1,36) = 53.90,$

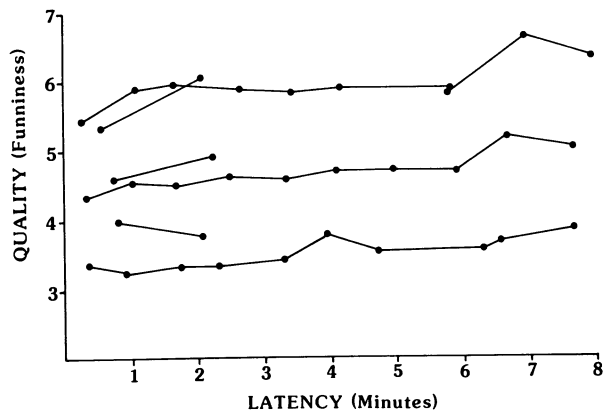


Figure 1. Humor rating (quality) as a function of order and rate of output (latency).

$p < .0001$ ) and pictures differed in rate of eliciting captions [ $F(2,72) = 4.62, p < .02$ ]. The only significant interaction for latency was between number of captions requested and the order of captions [ $F(1,36) = 6.04, p < .02$ ]. Pictures did not interact with conditions.

### Average Quality

When 10 captions were written, the rating of their humor increased progressively [ $F(9,162) = 2.50, p < .02$ ]. The particular picture did not have an effect on caption funniness [ $F(2,36) < 1.0$ ], and there was no interaction [ $F(18,324) = 1.10$ ]. When only two captions were written, however, no effects were significant. The second caption was not reliably better than the first [ $F(1,18) = 1.62$ ], even though the subjects' latencies suggested that they were being more self-critical and waiting for a better idea. Pictures also did not differ in caption funniness [ $F(2,36) < 1.0$ ], and the interaction was insignificant [ $F(2,36) = 1.20$ ].

In fact, the first 2 captions of 10 were very similar in quality to only 2 captions [ $F(1,36) = 1.78$ ]. Furthermore, the change from first caption to the second was not significant [ $F(1,36) = 2.55$ ]. The only significant relationship was between pictures and order, which interacted significantly [ $F(2,72) = 3.72, p < .05$ ]. One picture showed a decline in quality from first caption to second, whereas the other two showed an increase whether the subjects were writing 2 captions or 10. The delay in production for only 2 captions was not a result of a change in average quality.

Finally, as reported by Johnson, Parrott, and Stratton (1968), the average quality of the 2-captions condition would be expected to be better than the average quality of the 10-captions condition. If subjects were more selective for fewer captions, the average should be pulled down with greater output. Instead, the average quality was 4.69 for the 10 captions and 4.74 for the 2, which was not a significant difference [ $t(36) = 0.24$ ]. Even though subjects did show improvement with increased numbers of captions, they did not show evidence of discriminating overall quality of their output.

### Best and Worst

The average improvement in caption quality as a function of output order might have been a function of increased funniness of the best captions, improved rejection of the worst captions, or a mixture of both. In order to evaluate these possibilities, the best and worst captions for each subject at each output position were examined. Figure 1 also presents these functions.

Although the quality of the best captions increased almost parallel to the average, the change is not significant [ $F(9,162) = 1.35, p = .21$ ]. Nevertheless, the best caption of 10 was better than the best of 2 (8.24 vs. 6.37) [ $t(36) = 5.65, p < .0001$ ]. This finding does not correspond with that of Johnson et al. (1968), who found that "best" responses were equivalent regardless of how many

total responses were made; however, their tasks may have resulted in less divergent responses than did the present task. For example, their cartoon task presented four squares with a caption required for the last. Such context probably restricted the alternatives. Furthermore, Sheppard (1983) found photographs to be funnier than cartoons based on them.

In any case, the difference was a result of the extended output. The best of the first 2 captions of 10 were not significantly better than the best of 2 (6.53 vs. 6.37) [ $t(36) = 0.40$ ]. Although slower in rate, the subjects who only wrote 2 captions did not wait until they had thought up their potentially best responses.

As for an improvement in a quality threshold, the progressive rise for worst captions with an output of 10 (also shown in Figure 1) is not significant [ $F(9,123) = 1.28, p = .25$ ]. When compared with the worst of 2 captions, the worst of 10 is still worse (3.36 vs. 2.15) [ $t(36) = 4.98, p < .0001$ ]. In fact, this disadvantage also shows up in the first 2 captions of 10 (3.36 vs. 2.79) [ $t(36) = 2.15, p < .02$ ]. Apparently, the relative delay for subjects writing only 2 captions was caused by rejecting poor captions rather than by thinking up better ones. Furthermore, as shown by Derks (in press), a subject writing 10 captions produced his/her best caption after the worst 64.5% of the time, which is significantly different from the 50% expected if the 2 were both generated randomly [ $t(18) = 2.23, p < .02$ ].

## DISCUSSION

If humor is creative and funniness is the quality of that creation, then early ideas are not as creative as later ones. Furthermore, many ideas will produce more good ideas than will few, since "few" is determined by eliminating poor ideas rather than by selecting good ones. These results also suggest, but do not prove, that having poor ideas can lead to better ones.

The use of humor production to examine these issues is unusual. The task may have resulted in such divergent and idiosyncratic answers as to tell nothing about usual creativity and problem solving. It seems, however, that real creativity in the arts and sciences must be unique.

Therefore, the more the experimental task produces divergent responses, the more closely it will model art and invention.

## REFERENCES

- ARIETTI, S. (1976). *Creativity: The magic synthesis*. New York: Basic Books.
- BRODZINSKI, D. M., & RUBIEN, J. (1976). Humor production as a function of sex of subject, creativity, and cartoon content. *Journal of Consulting & Clinical Psychology*, **44**, 597-600.
- CHRISTENSEN, P. R., GUILFORD, J. P., & WILSON, R. C. (1957). Relations of creative responses to working time and instructions. *Journal of Experimental Psychology*, **53**, 82-88.
- COLELL, C. A., & DOMINO, G. (1980). Humor preferences and creativity. *Journal of Creative Behavior*, **14**, 215, 221.
- DERKS, P. L. (in press). Humor production: An examination of three models of creativity. *Journal of Creative Behavior*.
- DERKS, P. L., LEICHTMAN, H. M., & CARROLL, P. V. (1975). Production and judgment of "humor" by schizophrenics and college students. *Bulletin of the Psychonomic Society*, **6**, 300-302.
- DERKS, P. L., LEWIS, D. L., & WHITE, R. O., JR. (1981). A comparison of ratio and category scales of humour. *Canadian Journal of Behavioural Science*, **13**, 226-237.
- FRANKLIN, V. (1959). *Classics of the silent screen*. New York: Citadel.
- JOHNSON, D. M., PARROTT, G. L., & STRATTON, R. P. (1968). Production and judgment of solutions to five problems. *Journal of Educational Psychology, Monograph Supplement*, **59**(6, Pt. 2).
- KOESTLER, H. (1964). *The act of creation*. New York: Macmillan.
- MCGHEE, P. E. (1979). *Humor: Its origin and development*. San Francisco: W. H. Freeman.
- MEADOW, A., & PARNES, S. J. (1959). Evaluation of training in creative problem solving. *Journal of Applied Psychology*, **43**, 189-194.
- PARNES, S. J. (1961). Effects of extended effort in creative problem solving. *Journal on Educational Psychology*, **52**, 117-122.
- ROSIN, S. A., & CERBUS, G. (1984). Schizophrenics' and college students' preference for and judgment of schizophrenic versus normal humorous captions. *Journal of Psychology*, **118**, 189-195.
- SHEPPARD, A. (1983). Effect of mode of representation on visual humor. *Psychological Reports*, **52**, 299-305.
- TREADWELL, Y. (1970). Humor and creativity. *Psychological Reports*, **26**, 55-58.
- WEISBERG, R. W. (1986). *Creativity: Genius and other myths*. New York: W. H. Freeman.
- ZIV, A. (1984). *Personality and sense of humor*. New York: Springer.

(Manuscript received for publication June 1, 1987.)