



Brief article

Disfluency prompts analytic thinking—But not always greater accuracy: Response to Thompson et al. (2013)



Adam L. Alter^{a,*}, Daniel M. Oppenheimer^b, Nicholas Epley^c

^a Stern School of Business, New York University, 40 West 4th Street, Suite 811, New York, NY 10012, United States

^b Associate Professor of Marketing, UCLA Anderson School of Management, United States

^c John T. Keller Professor of Behavioral Sciences, University of Chicago, Booth School of Business, United States

ARTICLE INFO

Article history:

Received 29 October 2012

Revised 7 January 2013

Accepted 9 January 2013

Available online 16 February 2013

Keywords:

Fluency

Disfluency

Dual-system processing

Metacognition

Heuristics

Analytical thinking

System 1 processing

System 2 processing

ABSTRACT

In this issue of *Cognition*, Thompson and her colleagues challenge the results from a paper we published several years ago (Alter, Oppenheimer, Epley, & Eyre, 2007). That paper demonstrated that metacognitive difficulty or disfluency can trigger more analytical thinking as measured by accuracy on several reasoning tasks. In their experiments, Thompson et al. find evidence that people process information more deeply—but not necessarily more accurately—when they experience disfluency. These results are consistent with our original theorizing, but the authors misinterpret it as counter-evidence because they suggest that accuracy (and even confidence) is a measure of deeper processing rather than a contingent outcome of such processing. We further suggest that Thompson et al. err when they discriminate between “perceptual fluency” and “answer fluency,” the former of which is an element of the latter. Thompson et al. advance research by adding reaction time as a measure of deeper cognitive processing, but we caution against misinterpreting the meaning of accuracy.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

This issue of *Cognition* features a paper by Thompson et al. (2013) that both complements and challenges a paper we published several years ago (Alter, Oppenheimer, Epley, & Eyre, 2007). We're pleased that Thompson et al. engaged with our work, but also eager to address two areas of disagreement. First, Thompson et al. suggest that their findings fail to replicate ours. In fact, they confirm our account except when the questions are too difficult for participants to answer correctly. Both papers ultimately arrive at the same conclusion. Second, Thompson et al. discriminate between two forms of fluency—“perceptual fluency” and “answer fluency”—that we think are no more independent than apples are from fruit. Instead, as the analogy suggests, perceptual fluency is part of the ease with which an answer comes to mind. Since perceptual fluency

corresponds to a genuine cognitive process, whereas answer fluency incorporates a range of operations, we question the value of comparing their relative contributions to judgment.

Our paper (Alter et al., 2007) consisted of four studies in which participants were less confident, and consequently more thoughtful, about their responses to various questions when they experienced disfluency—the subjective experience of difficulty while processing information. In one experiment, for example, participants were more likely to reconsider intuitively appealing but incorrect responses to three problems (the Cognitive Reflection Test, or CRT: Frederick, 2005) when the questions were printed in a difficult-to-read font. We theorized that participants misattributed the difficulty of reading the material for the difficulty of the problems themselves, thereby encouraging them to think more carefully. This careful thinking then improved performance on analytical reasoning tasks.

This basic effect of disfluency on more analytical processing has been replicated, extended, and applied a

* Corresponding author. Tel.: +1 212 998 0142.

E-mail address: aalter@stern.nyu.edu (A.L. Alter).

number of times since our original demonstration (e.g., Cokely, Parpart, & Schooler, 2009; Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Kappes & Alter, 2013; Oppenheimer & Alter, *in press*; Song & Schwarz, 2008; also see Simmons & Nelson, 2006, Study 13, for a relevant precursor). Like any scientists, we are pleased that Thompson et al. found the results important enough to examine further. Their approach is particularly appealing because it includes reaction times that serve as a more direct measure for analytical reasoning. Although other researchers have suggested that participants process information more analytically when they experience disfluency, Thompson et al. provide the first direct demonstration, to our knowledge, that disfluency prompts people to spend more time considering their responses. We agree with Thompson et al.'s suggestion that researchers should measure response times as well as accuracy rates to capture processing depth, and we believe this insight is the paper's primary contribution.

An alternative experimental approach (that we adopted in our experiments) is to choose tasks that inspire categorically different responses depending on whether participants are relying on intuitive and analytical thinking. This approach measures the depth of cognitive processing indirectly, depending on which category of response a participant exhibits. To be a valid measure of cognitive processing, however, this approach requires that the different types of cognitive processing actually lead to categorically different responses. If the items are so difficult that no amount of careful thinking will lead to a correct (or different) response, then a person's answer cannot tell us about her style of thinking.

Correct responses are therefore an inappropriate proxy for deeper processing when the reasoning tasks are either too difficult (and impossible to answer no matter much a person thinks about them), or too easy (and capable of being answered correctly based on intuition alone; for more on this point see Alter et al., 2007, Experiment 4). We believe that using overly difficult tasks to measure cognitive processing is precisely the reason for the apparent discrepancy between Thompson et al.'s results and our original results, and their paper consequently highlights the importance of calibrating cognitive tasks before using them to assess differences in processing depth.

2. Reconciling Thompson et al. (2013) and Alter et al. (2007)

Despite the differences in confidence and accuracy in Thompson et al. (2013) and Alter et al. (2007), both papers demonstrate the more fundamental principle that people process information more deeply when they experience cognitive disfluency. The differences arise because Thompson et al. find only limited evidence that disfluency improves accuracy. However, disfluency can encourage deeper processing without improving accuracy rates when the task is simply too difficult for participants to answer. Our theory proposed that disfluency triggers deeper cognitive processing, but deeper cognitive processing does not ensure accuracy in and of itself. Ten-year-olds are unable to solve calculus problems no matter how hard they think,

and would be unable to solve them even if they are told directly that the problems are difficult and require considerable mental effort. The participants in Thompson et al.'s studies are certainly more sophisticated than ten-year-olds, but in both cases the questions are too difficult to uncover a link between processing depth and accuracy. The relationship between mental effort and mental performance is therefore similar to the relationship between physical effort and physical performance. None of the authors on this comment, for instance, can dunk a basketball no matter how hard he tries.

Thompson et al.'s results demonstrate that participants indeed thought more carefully about the tasks at hand, but were not always capable of providing the right answer even with more careful thought. In Experiments 1b, 2a, and 2b, participants spent more time generating an answer when they adopted a disfluent facial expression (Experiment 1b: $F[1, 70] = 8.27, p = .005, \eta_p^2 = .11$), or when the problems were presented in a disfluent font (Experiment 2a: $F[1, 46] = 10.86, p = .002, \eta_p^2 = .06$; Experiment 2b: $F[1, 40] = 4.07, p = .05, \eta_p^2 = .09$). As our theory would predict, disfluency triggers more careful and analytical thinking.

So why, if people thought more deeply, were their answers no more accurate than the answers of their peers in the studies' fluent conditions? The answer looks to be the same reason that none of us can dunk a basketball. Participants found the tasks to be too difficult even when they were explicitly instructed to think more carefully. For example, in Experiment 1a participants completed a series of syllogistic reasoning problems, first providing an intuitive response, and then thinking more carefully before providing a more deeply considered follow-up response. All responses were binary, requiring participants to indicate whether or not a conclusion followed logically from a preceding statement. Participants who responded randomly would therefore have answered approximately 50% of the questions correctly. As Thompson et al. reported in Table 1 of their manuscript, participants answered between 53% and 54% of the questions correctly when they relied on their intuitions, and only 55% of the questions correctly when they reconsidered their initial responses and thought about the problems more deeply. These overall accuracy rates are barely better than one would expect by chance alone, suggesting that the tasks were simply too difficult for additional thinking to improve performance. Indeed, the clearest way to identify a task as "difficult" is to show that overall accuracy rates are no better than chance.

In contrast to the results in Experiment 1a in Thompson et al., participants in Experiment 4 of Alter et al. chose their responses from among nine options, and managed to answer an average of 54% of the questions correctly across the fluent and disfluent conditions. These accuracy rates were therefore 43% higher than the 11% rate expected from chance alone, suggesting that their responses reflected a degree of skill. Because even a direct instruction to think harder failed to improve accuracy rates in Thompson et al.'s Experiment 1a, the questions were simply too difficult to serve the purpose for which they were chosen: to distinguish between shallower and deeper processing in the tested sample of students. Likewise, participants in

Experiment 1c struggled with the Cognitive Reflection Test (CRT). The CRT is notoriously difficult, and countless norming experiments have shown that only students from elite universities routinely score an average of more than two of the three questions correctly (Frederick, 2005). Consistent with this finding, when Thompson et al. presented the CRT to a large, representative sample of online participants, and students at a large Canadian public university, they answered an average of between .78 and 1.26 of the three questions correctly. As Thompson et al. note, the participants in our experiments were from elite universities who scored considerably higher on this test.

But the strongest evidence for the consistency between Thompson et al.'s results and Alter et al. (2007) come from the final two experiments, 3a and 3b. In these experiments, disfluency improved CRT scores among students with high SAT and IQ scores but not those with lower SAT & IQ scores. In contrast to their peers, and like the Princeton and Harvard students in our paper (Alter et al., 2007, Experiment 1), those high-performing students were able to answer the CRT questions correctly when they recruited additional cognitive resources (see also Cokely et al., 2009, for a similar result). Far from contradicting our paper, the results in Experiments 3a and 3b are consistent with ours.

The practical question, then, is how researchers should decide, *a priori*, whether a task is appropriate for experiments like these. Indeed, like Thompson et al., we encountered a similar problem in Experiment 4 of our manuscript, when participants were incapable of answering the two most difficult of six syllogism problems. As a result, we excluded those two problems from our analyses because accuracy is a poor measure of processing when questions are too difficult (Alter et al., 2007, p. 574).

To avoid these issues, we suggest a simple test. Before deciding that a task is appropriate, researchers should run a pilot experiment with two groups of participants. The first group should be told, explicitly, that the task is difficult and requires considerable mental effort—a heavy-handed proxy for more subtle disfluency manipulations. The second group should complete the task without those instructions. If the former group outperforms the latter, then deeper processing improves accuracy on that task. If not, then the task is either too difficult for more careful thinking to make any difference, or too easy so that more careful thinking is not needed to generate the right answer. Consistent with the recent push toward improving research practices in psychological science (e.g., Simmons, Nelson, & Simonsohn, 2011), researchers should decide *a priori* how many participants will complete their pilot experiment, and how they will distinguish appropriate from inappropriate tasks based on the data that emerge from that pre-test.

Overall, the seven experiments in Thompson et al. (2013) are consistent with the four experiments in our manuscript (Alter et al., 2007), despite the authors' interpretation to the contrary. When the tasks are properly calibrated, as they were for the high-performing students in Experiments 3a and 3b, disfluency prompts deeper processing which in turn improves participants' accuracy rates. In contrast, when the tasks are too difficult disfluency encourages deeper processing without improving

accuracy rates (Experiments 1b, 2a, and 2b). Thompson et al. argue that their results are inconsistent with ours. In fact, we believe the pattern of their results strongly supports our theory, and is consistent with the evidence we presented.

3. Questioning “Answer Fluency”

In his classic book, *On the Concept of Mind*, philosopher Gilbert Ryle (1949) described the case of a visitor to Oxford University who looked at the buildings and library, and then asked “But where is Oxford?” The man's mistake lay in a class inclusion error: his failure to recognize that the buildings and library themselves constituted the superordinate concept of Oxford University. The same illogic underlies Thompson et al.'s distinction between “answer fluency” and “perceptual fluency.” One of the central aims of their paper is to measure and compare how strongly these two forms of fluency shape judgment.

Thompson et al. define “answer fluency” as the ease with which a response is generated—a concept that we and many other researchers in the field simply call “fluency” (for reviews see Alter & Oppenheimer, 2009; Schwarz, 2004). As we noted in our review, answers come to mind more easily for numerous reasons: because the answer to the question was discovered very recently (retrieval fluency); because the respondent happened to be pondering a related topic (priming fluency); because the question was phrased simply (linguistic fluency); or because the question was printed in a clearer font (perceptual fluency). Each of these forms of fluency corresponds to a particular cognitive operation. For example, we decomposed memory-based fluency into encoding fluency and retrieval fluency, which correspond to the cognitive operations of encoding and later retrieving information from memory. We similarly divided perceptual fluency into visual perceptual fluency and auditory perceptual fluency, which parallel the processes of vision and audition. Along with numerous other instantiations of fluency (see Fig. 1 in Alter & Oppenheimer, 2009), the combined ease with which people accomplish these cognitive tasks forms a global sense of whether the question was answered with ease (fluently) or with difficulty (disfluently). One classic illustration is a study by Reber and Schwarz (1999), in which participants believed that trivia responses were more likely to be true when they were presented more clearly. Perceptual fluency—the sense of ease associated with perceiving the trivia questions and responses—imbued those responses with a sense of truth, familiarity, or rightness. Here perceptual fluency fed directly into what Thompson et al. call “answer fluency.”

Because perceptual fluency is just one of many inputs that comprise the more general experience of fluency (or “answer fluency” to use Thompson et al.'s label), it should come as no surprise that the superordinate category explains more variance in participants' responses than the subordinate category does. It makes no more sense to compare the relative impact of perceptual and answer fluency on judgments than it does to compare the relative impact of Big Bird and muppets in general on children's enjoyment of Sesame Street.

Finally, it is worth noting that, while Thompson et al. try to draw a strong distinction between their theoretical position and ours, their suggestion that we had focused on ‘perceptual fluency’ is a misrepresentation of our position. In our original paper, our primary construct was “metacognitive difficulty” (pg. 569), which is virtually indistinguishable from Thompson et al.’s “answer fluency”. In fact we only used the phrase “perceptual fluency” once in the entire paper: in the reference section.

4. Conclusion

Despite the surface differences between our results (Alter et al., 2007) and those of Thompson et al. (2013), both papers suggest the same conclusion: that people respond to metacognitive difficulty by recruiting additional mental resources. Their responses improve when the task isn’t too difficult or so easy that mental engagement doesn’t provide an advantage, but sometimes—as Thompson and her colleagues demonstrated—accuracy is a poor proxy for deeper thought when the tasks at hand are poorly calibrated. Moreover, Thompson et al. (2013) attempt to draw a distinction between our theoretical positions which does not exist. Their paper provides strong, convergent evidence for our original hypothesis, rather than a competing account.

References

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–576.
- Cokely, E. T., Parpart, P., & Schooler, L. J. (2009). On the link between cognitive control and heuristic processes. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society* (pp. 2926–2931). Austin, TX: Cognitive Science Society.
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold: Effects of disfluency on educational outcomes. *Cognition*, 118, 111–115.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.
- Kappes, H. B., & Alter, A. L. (2013). *Disfluency enhances fantasy*. New York University. Unpublished manuscript.
- Oppenheimer, D. M., & Alter, A. L. (in press). Disfluency sleeper effect: Disfluency today promotes fluency tomorrow. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour*. London: Psychology Press.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8, 338–342.
- Ryle, G. (1949). *The concept of mind*. Chicago, IL: University of Chicago Press.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14, 332–348.
- Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence. Choosing between intuitive and non-intuitive alternatives. *Journal of Experimental Psychology: General*, 135, 409–428.
- Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, 26, 791–799.
- Thompson, V., Prowse Turner, J., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*. [Page and issue details to be populated upon publication.]