Update

Research Focus

Seduction without cause: uncovering explanatory neurophilia

J.D. Trout

Philosophy Department and the Parmly Sensory Sciences Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, IL 60626, USA

Credibility is a cherished currency in science, but its cues can be counterfeit. A novel series of experiments by Weisberg and her colleagues show that non-expert consumers of behavioral explanations assign greater standing to explanations that contain neuroscientific details, even if these details provide no additional explanatory value. Here, we discuss the part that this 'placebic' information might play in producing a potentially misleading sense of intellectual fluency and, consequently, an unreliable sense of understanding.

People often accept inaccurate explanations because they feel right [1,2]. In psychology, cognitive neuroscience explanations of human behavior induce this glow of recognition and convey the blissful aha! Thanks to heavy media attention, the lay public can now imagine Broca's area lighting up as we work to find the right word and envision dark expanses in the prefrontal cortex of a remorseless criminal. The media report that scientists can read a magnetic resonance imaging scan of the visual system and identify the scene that the subject has just viewed, highlighting the vivid prospects for transforming daily life. We are then invited to speculate whether a computer could read a quadriplegic's brain scan and execute their commands and whether scan-reading could provide the blueprint for all-purpose brain-reading devices that escort us through a person's visual experiences or even their dreams [3]. These tidy, intuitive and vividly imaginable findings of cognitive neuroscience are gripping. Yet, new research by Weisberg *et al.* [4] gives us reason to question the source of this eager embrace.

Weisberg and her colleagues explain this seductive but knowledge-eroding effect in terms of the potency of irrelevant or 'placebic' information. Placebic information has a peculiar effect on the reasoner. This extraneous information, especially neuroscientific information, gives people a mistaken feeling that they have received a good explanation.

But why would placebic information (rather than, say, false or shocking information) create a sense of fluency? This question goes beyond the scope of Weisberg *et al.* [4], but the conceptual connections are easy to trace. Placebic information has characteristics that promote the feeling of intellectual fluency. Fluent processes are fast, coherent and automatic. As with many cases of conceptual priming, the technical vocabulary and causal taxonomy of placebic neuroscientific information might activate conceptual representations contained in psychological and neuroscientific explanations. Thus, even if information is irrelevant to the content of the explanation, it might be quite relevant to the good feeling of fluency experienced when we assemble and process an explanation [5].

Good and bad explanations for behavior

In philosophy, great debates rage about the requirements of good explanation. The nature of good explanation might not be revealed anytime soon, but Weisberg *et al.* [4] easily sidestep this problem by creating stimulus materials that everyone agrees are bad explanations because they violate a formal constraint on explanation – they are circular. They do this to ask whether people are persuaded by cognitive neuroscience explanations of behavior and whether they are persuaded for good reasons, in addition to testing a potentially embarrassing hypothesis: maybe the mere mention of a neuroscientific detail is enough to credential an explanation. This hypothesis can be tested by appealing to neuroscientific information that has no relevance to the quality of the explanation. Weisberg *et al.* [4] show that people prefer explanations that contain merely placebic neuroscientific information – information irrelevant to the explanation's quality – over those that do not.

Their experiments cross quality of explanation (good versus bad) with presence of neuroscientific information (with versus without). The 'bad', circular explanations conveyed information that is logically irrelevant to the truth of the explanation. Overall, non-experts found good explanations significantly more satisfying than bad explanations, and explanations with neuroscientific information more satisfying than those without. In addition, bad explanations with neuroscientific detail enjoyed a neurophilic premium; neuroscientific information produced a special boost in perceived accuracy.

Most people would like to think that a little training in cognitive neuroscience would defeat such seduction. But training in critically appraising behavioral explanations does not help much – similar to novices, students enrolled in an introductory neuroscience class appraised explanations with neuroscientific information as more plausible than those without. Furthermore, bad explanations benefited more from placebic neuroscientific information than did good explanations.

The main effect of experiment three by Weisberg *et al.* [4] on experts emerged not in recognizing how bad the non-neuroscientific explanations were but in being less

Corresponding author: Trout, J.D. (jtrout@luc.edu).

impressed with the good neuroscientific explanations. There are similar effects on experts in other domains, in which experts moderate their assessments. Experts might be less hoodwinked by neuroscientific explanations for the same reason that improving a person's skills reduces their overconfidence about them [6].

Except for truly expert subjects, placebic neuroscientific-explanation-primed conceptual vehicles that deliver feelings of intellectual satisfaction. And placebic information is not the sole proprietor of fluency; this sense of understanding is conveyed by several well-documented psychological effects, such as the feeling of knowing, illusion of explanatory depth and tip-of-the-tongue experiences [7].

Truth and consequences

This neurophilic fluency flourishes wherever heuristics in psychology are reductionist. Reductionism, as a trend in science, is the effort to locate causes internal to systems and to identify the causal powers of macroscopic phenomena with their molecular constituents. Cast psychologically, reductionism acts as a heuristic for complicated problems; by focusing on what is usually a small number of local causal factors, it renders a complex domain cognitively tractable [8]. McCabe and Castel [9] found that participants who read the same articles on cognitive neuroscience research were more persuaded if it was accompanied by a brain image, when compared with a topographical map of brain activation, bar graphs or no image at all. The brain image is more vivid, more easily visualized. And so this effect might be rooted in the early findings in heuristics and biases tradition, in which fluency and tractability meet [10]. One reason the Weisberg *et al.* [4] findings are especially striking is that they produce the same misleading fluency via a more verbal, less visualizeable medium.

Neuroscientific details qualify, and sometimes replace, distinctly cognitive or conceptual features of behavioral explanations. That trend might seem desirable because modern neurophilia promises to make manageable the internal components that explain behavior. The promise of cognitive tractability enhances fluency, but not necessarily accuracy. All too often humans interpret the positive hedonic experience of fluency as a mark of genuine understanding. Thus, Weisberg *et al.*'s [4] findings should urge us to ask analogous questions about other patterns of scientific explanation. Would biologists be more satisfied with bad chemical explanations for a biological phenomenon or chemists with bad physics explanations of a chemical process?

Among philosophers, the usefulness of reductionist explanations is an open empirical question. Historical evidence for reduction is uneven. Thermodynamics might disappear into statistical mechanics, or light into radiation – but just as often domains stubbornly resist reduction. And with good reason: reductionist explanations of complex systems such as bat audition, the ontogenesis of human language or lake ecology might misrepresent their sensitivity to changes [11]. But whatever the actual attractions of reductionism, psychology highlights its perceived usefulness. The psychology of science might explore the generality of this reductionist impulse. For example, do people favor biological explanations that contain placebic chemical information or chemical explanations that mention extraneous physics concepts?

The public's neurophilia is worth studying in part because it might be exploited – prematurely, even cynically – by public-policy entrepreneurs attempting to shape cultural tastes and political convictions. Such sculpting might begin with irrelevant placebic information and pick up momentum by introducing a seductive fluency. Because fluency rides upon psychological habit and the regimented products of motor learning, voluntary programs of education or cautionary directives are unlikely to douse this seduction.

Meanwhile, we can at least appreciate that neuroscience is not the only source of temptation in theory construction. Do intelligent design explanations enjoy enhanced respectability, however spurious, because they use words such as 'immune system', '*Escherichia coli* flagellum' and 'blood clotting'? It would also be useful to know whether a groundless neurophilia bolsters other kinds of explanations, such as functional explanations (of the form 'the function of X is to Y'; see Ref. [12]). In the near future, issues of great human significance will come before the American electorate – stem-cell therapies, genetic engineering and imaging techniques bearing on legal competence, to name just a few. Weisberg *et al.* [4] has edged us toward understanding the causes of uncritical assessments of scientific explanations.

References

1 Trout, J.D. (2002) Scientific explanation and the sense of understanding. *Philos. Sci.* 69, 212–233
2 Trout, J.D. (2007) The psychology of scientific explanation. *Philosophy Compass* 2, 564–591
3 Haynes, J.D. and Rees, G. (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* 15, 1301–1307
4 Weisberg, D.S. *et al.* (2008) The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* 20, 470–477
5 Winkielman, P. *et al.* (2003) The hedonic marking of processing fluency: implications for evaluative judgment. In *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (Musch, J. and Klauer, K.C., eds), pp. 189–217, Lawrence Erlbaum Associates
6 Kruger, J. and Dunning, D. (1999) Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134
7 Keil, F.C. (2006) Explanation and understanding. *Annu. Rev. Psychol.* 57, 227–254
8 Wimsatt, W. (2007) *Re-Engineering Philosophy for Limited Beings,* University of Chicago Press
9 McCabe, D.P. and Castel, A.D. (2008) Seeing is believing: the effect of brain images on judgments of scientific reasoning. *Cognition* 107, 343–352
10 Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: Heuristics and biases. *Science* 185, 1124–1130
11 Garfinkel, A. (1981) *Forms of Explanation,* Yale University Press
12 Lombrozo, T. and Carey, S. (2006) Functional explanation and the function of explanation. *Cognition* 99, 167–204

1364-6613/$ – see front matter
doi:10.1016/j.tics.2008.05.004 Available online 5 July 2008