

Deconstructing the seductive allure of neuroscience explanations

Deena Skolnick Weisberg* Jordan C. V. Taylor† Emily J. Hopkins†

Abstract

Previous work showed that people find explanations more satisfying when they contain irrelevant neuroscience information. The current studies investigate why this effect happens. In Study 1 ($N=322$), subjects judged psychology explanations that did or did not contain irrelevant neuroscience information. Longer explanations were judged more satisfying, as were explanations containing neuroscience information, but these two factors made independent contributions. In Study 2 ($N=255$), subjects directly compared good and bad explanations. Subjects were generally successful at selecting the good explanation except when the bad explanation contained neuroscience and the good one did not. Study 3 ($N=159$) tested whether neuroscience jargon was necessary for the effect, or whether it would obtain with any reference to the brain. Responses to these two conditions did not differ. These results confirm that neuroscience information exerts a seductive effect on people's judgments, which may explain the appeal of neuroscience information within the public sphere.

Keywords: explanation, neuroscience, reasoning, seductive allure.

1 Introduction

Attention to neuroscience is growing within the public sphere. Neuroscientific findings now play a key role in public conversations about economics, marketing, and the law, among other areas (e.g., Ariely & Berns, 2010; Camerer, Loewenstein & Prelec, 2005; Farah, 2012; Greene & Cohen, 2004; Roskies, 2002; Satel & Lilienfeld, 2013). For example, neuroscience data are often used in courtrooms as evidence of a defendant's responsibility or guilt (Morse, 2011; Saks, Schweitzer, Aharoni & Kiehl, 2014; Schweitzer et al., 2011). But it is not entirely clear how members of the public view these findings. Do they understand the role that neuroscience information plays in explanations of people's beliefs and behaviors?

Previous research suggests that the answer to this question is “no”. People are unduly swayed to think favorably of psychology explanations that include references to neuroscience—even when such neuroscience information is logically irrelevant to the explanations (Weisberg, Keil, Goodstein, Rawson & Gray, 2008). In this study, subjects read descriptions of psychological phenomena. Each phenomenon was followed by one of four types of explanation, constructed by crossing explanation quality (good or bad) with neuroscience information (present or absent). Cru-

cially, the neuroscience information was irrelevant to the logic of the explanations and even made the good explanations worse, according to the ratings of experts.

When the explanations contained neuroscience information, ratings were significantly higher than when they did not. This was especially true for the bad explanations (perhaps because people have trouble detecting circularity in arguments, see Rips, 2002). That is, non-experts judged that psychological phenomena are explained better using the language of neuroscience, although this language should make no difference, assuming that an explanation's quality is drawn primarily from the strength of its logic. One recent study (Scurich & Shniderman, 2014) also found that subjects give higher ratings to studies that included neuroscience information, but only when the conclusions of these studies confirmed their prior beliefs. However, the absence of a no-neuroscience control condition in this study makes it difficult to draw firm conclusions about the general effect of neuroscience information.

Three other studies did include the appropriate controls, and both confirmed Weisberg et al.'s (2008) findings. One used the same stimuli in an exact replication (Fernandez-Duque, Evans, Christian & Hodges, 2015). The other two used different sets of stimuli in a conceptual replication (Rhodes, Rodriguez & Shah, 2014; Rhodes & Shah, 2015), in which subjects read a mock news article describing psychological research; the article either did or did not contain irrelevant neuroscience information. Neuroscience information thus exerts a *seductive allure effect*, whereby people without advanced training believe that references to brain processes improve the quality of a psychological explanation, even when these references are logically irrelevant. This effect could be thought of as part of a family of

We thank Matthew Bateman, Martha Farah, Diego Fernandez-Duque, Frank Keil, and the members of the Penn Cognition & Development Lab. This research was funded by the Templeton Foundation (Varieties of Understanding project grant to DSW).

Copyright: © 2015. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*University of Pennsylvania, Department of Psychology, 3720 Walnut St., Solomon Labs, Philadelphia, PA 19104. Email: deena.weisberg@psych.upenn.edu.

†University of Pennsylvania.

heuristics that people use for judging the quality of explanations, which includes teleological information (Lombrozo & Carey, 2006) and an intuitive sense of satisfaction (Trout, 2002).

One study claimed that neuroscience images are responsible for the effect (McCabe & Castel, 2008), suggesting that people are seduced by the visual appeal of images generated by fMRI scans and other neuroscientific techniques. However, many later studies have failed to replicate this finding (Gruber & Dickerson, 2012; Hook & Farah, 2013; Keehner, Mayberry & Fischer, 2011; Michael, Newman, Vuorre, Cumming & Garry, 2013; see Farah & Hook, 2013, for review). To test directly whether brain images add value to explanations that already contained neuroscience text, Fernandez-Duque et al. (2015) presented subjects with explanations that either contained no neuroscience information, contained irrelevant neuroscience information, or contained irrelevant neuroscience information and were accompanied by a neuroscience image. These researchers found that people rated explanations with neuroscience information as better than explanations without this information, as noted above, but images did not have any additional effect. Further, Weisberg et al. (2008), Fernandez-Duque et al. (2015), Rhodes et al. (2014), and Rhodes and Shah (2015) obtained the seductive allure effect without the use of any pictures. These studies strongly suggest that neuroscience imagery is not the source of the effect.

Why, then, does this effect happen? The importance of answering this question becomes evident when we examine the many ways in which neuroscience information is used (and misused) in the public sphere. The proliferation of headlines proclaiming that some drug or activity “literally changes your brain” illustrates both how appealing neuroscience information is to the general public and how poorly this information is understood. To take a weightier example, attorneys may appeal to neuroscience-based evidence in order to convince a jury of a legal fact. But because this kind of information is intuitively compelling even when it is irrelevant, such evidence may unduly bias the jury, potentially threatening the fairness of the judicial system (see Greene & Cohen, 2004; Morse, 2004). Similarly, in the field of education, unsubstantiated claims about how children’s brains change or fundamental differences between boys’ and girls’ brains can lead to the implementation of educational policies or practices that seem appealing but may not actually benefit the students (see Bruer, 1997; Goswami, 2006).

Learning why neuroscience information is alluring can help us to develop techniques to reverse some of these trends. The current studies begin to address this issue by investigating three factors that might contribute to the seductive allure effect: length (Study 1), explicit appeal of neuroscience (Study 2), and jargon (Study 3). In terms of length, the explanations in Weisberg et al. (2008) that contained irrelevant neuroscientific information were always longer

than the explanations that did not. Subjects may have simply rated longer explanations as better. Indeed, other work showed that people prefer longer explanations, even if the added length did not add to the explanation’s quality (Kikas, 2003). Study 1 thus begins our investigation of this effect by replicating Weisberg et al. (2008) with the addition of a control for the length of the explanations.

A second possible explanation for the effect is that neuroscience information may appeal due to its authoritative aesthetic: Explanations containing neuroscience information may look as though they have come from a suitably scientific process, and so may be perceived as trustworthy and therefore convincing, regardless of their content (see Sperber, 2010). We address this issue in Study 2 by asking subjects to directly compare good and bad explanations when they do and do not contain neuroscience information.

The third possibility that we investigate is that people are attracted to any kind of scientific-sounding jargon because they believe that use of these fancy terms signals higher-quality science. Indeed, math-based jargon has precisely this effect (Eriksson, 2012). We address this issue in Study 3 by comparing subjects’ ratings of explanations that use simple references to brain processes with their ratings of explanations that use more technical terms.

2 Study 1

Study 1 was designed to determine whether the seductive allure effect results from subjects’ responses to neuroscience information itself or from the tendency for explanations containing neuroscience information to be longer than explanations without this information. Previous work suggests that length does not account for the effect: Fernandez-Duque et al. (2015) found that explanations with added neuroscience information were rated more highly than unembellished explanations, but explanations with added social psychology information were not. In addition, Rhodes et al. (2014) found that stimuli with neuroscience information were rated more highly than length-matched stimuli without neuroscience information. These results suggest that the seductive allure effect cannot be accounted for solely by the explanations’ length.

Study 1 continued this investigation of the role of length and addressed a potential issue with the method used in previous studies. Both Fernandez-Duque et al. (2015) and Rhodes et al. (2014) controlled for length by making the without-neuroscience explanations longer, so as to match the length of the with-neuroscience stimuli. But this additional information may have affected how subjects rated the without-neuroscience explanations. For example, Fernandez-Duque et al. (2015) compared explanations with superfluous information from social science or hard sciences to those with superfluous neuroscience information. How-

ever, this information from other fields may have seemed less relevant to the explanations than the neuroscience information, potentially lowering subjects' ratings. Thus this design does not separate the effect of length from the effect of different types of added information. The current study made the with-neuroscience explanations shorter so as to match the length of the without-neuroscience stimuli. This more fully un-confounds the variables of length and neuroscience information.

Study 1 thus provides a more complete investigation of the potential effect of length on the seductive allure effect, which will allow us to determine how neuroscience information affects people's judgments. If the seductive allure effect is only due to a general tendency to judge longer explanations as better, then it should disappear when the explanations that do and do not contain neuroscience are matched for length. But if something about neuroscience information leads to more positive judgments of explanations, then the effect of neuroscience should remain regardless of the length of the explanation.

2.1 Method

Subjects. We recruited subjects from two populations: undergraduate students from the psychology subject pool at the University of Pennsylvania and workers on Mechanical Turk. Because previous work on this topic has primarily used undergraduates as subjects, we added the MTurk workers in order to assess the generality of the effect in a more representative population. This study included 204 undergraduates (143 women, 61 men; mean age = 19.8 years, range = 18–50) and 177 MTurk workers (85 women, 92 men; mean age = 37.5 years, range = 19–70). Undergraduates received course credit for participating in the study, and MTurk workers received 20 cents.

Design. Subjects were divided into 4 conditions according to a 2 (Neuroscience: with, without) x 2 (Length: long, short) design. These were both between-subjects variables, so an individual subject saw explanations that either all included or all did not include neuroscience information, and their explanations would all come from the same length category. There were 43 MTurk workers and 44 undergraduates in With Neuroscience-Long, 40 MTurk workers and 65 undergraduates in With Neuroscience-Short, 49 MTurk workers and 50 undergraduates in Without Neuroscience-Long, and 45 MTurk workers and 45 undergraduates in Without Neuroscience-Short. Quality was a within-subjects variable; for each trial, the survey software randomly determined whether to show the good or bad version of the explanation.¹

¹Due to the randomization, there were 47 subjects who saw either good explanations on every trial or bad explanations on every trial. The inclusion of these subjects did not affect any analyses, so they were left in the sample.

Materials. We selected four of the 18 items presented to subjects in Weisberg et al. (2008) and Fernandez-Duque et al. (2015) (babies' abilities to do simple arithmetic, attentional blink, gender differences in spatial reasoning, and differences between seeing and imagining objects; see supplemental materials for full stimulus items). These were items for which subjects in a pilot sample consistently judged the bad version of the without-neuroscience explanation as worse than the good version of that explanation. Each of the four items consisted of a description of a psychological phenomenon and eight different explanations for that phenomenon. The good explanations are the ones that the researchers themselves provided for the phenomena or that were provided in psychology textbooks. The bad explanations were circular restatements of the phenomena with no mechanistic information that could give a reason for the phenomenon. Items in all studies are in the supplement.

The explanations used in the Without Neuroscience-Short and the With Neuroscience-Long conditions exactly matched those used in Weisberg et al. (2008). To construct the Without Neuroscience-Long explanations, we added superfluous wording to the existing Without Neuroscience-Short explanations to make them the same length as the corresponding With Neuroscience-Long explanations. Importantly, this additional wording referred only to psychological constructs and never to other sciences, mathematics, or neuroscience, and this information did not add any value to the explanation. To construct the With Neuroscience-Short explanations, we edited the existing With Neuroscience-Long explanations to make them the same length as the corresponding Without Neuroscience-Short explanations. Regardless of length, the irrelevant neuroscience information was identical across the good and bad versions of the explanations for each phenomenon.

Procedure. All subjects completed an online survey distributed on Qualtrics. In each trial, subjects read a description of a psychological phenomenon, which appeared in isolation on the screen for 10 seconds before they were allowed to advance to the next screen. On the second screen of each trial, the phenomenon appeared again at the top, followed by one of the eight possible explanations for that phenomenon. Subjects were asked to judge how satisfying they found this explanation on a seven-point scale, from -3 (very unsatisfying) to +3 (very satisfying), with 0 as the neutral midpoint.

Each subject saw all four stimulus items, one per trial, in a randomized order. On each trial, the phenomenon was presented along with one version of the explanation. Subjects' condition determined whether they saw a long or short version and a with- or without-neuroscience version. Whether they saw a good or bad version of the explanation was randomly determined on each trial (as described above in the Design section). At the end of the survey, subjects provided basic demographic information: age in years, gender, and

Table 1: Study 1 mixed-effects linear regression model (* $p < .05$).

Predictor	Estimate	[95% CI]	t
Intercept	0.24	[0.14, 0.34]	5.00*
Item 2	0.24	[0.10, 0.38]	3.21*
Item 3	-0.61	[-0.74, -0.47]	-8.82*
Item 4	0.65	[0.21, 0.79]	9.41*
Length	0.12	[0.02, 0.21]	2.55*
Neuroscience	0.13	[0.03, 0.22]	2.70*
Group	0.15	[0.06, 0.25]	3.18*
Quality	0.36	[0.28, 0.44]	8.47*
Neuroscience x Item 2	-0.05	[-0.19, 0.10]	-0.72
Neuroscience x Item 3	0.20	[0.06, 0.34]	2.93*
Neuroscience x Item 4	-0.24	[-0.37, -0.10]	-3.45*
Group x Item 2	0.03	[-0.11, 0.17]	0.50
Group x Item 3	0.16	[0.01, 0.30]	2.34*
Group x Item 4	-0.19	[-0.34, -0.05]	-2.71*
Quality x Item 2	0.08	[-0.08, 0.20]	1.06
Quality x Item 3	-0.14	[-0.28, -0.02]	-1.94*
Quality x Item 4	-0.25	[-0.38, -0.12]	-3.57*

Note: This regression predicted subjects' ratings of the quality of the explanations. The intercept represents the Without Neuroscience condition, short explanations, undergraduate subjects, and bad explanations. Item is deviation coded, such that the coefficient for each level represents deviation from the grand mean; Item 1 is the reference level.

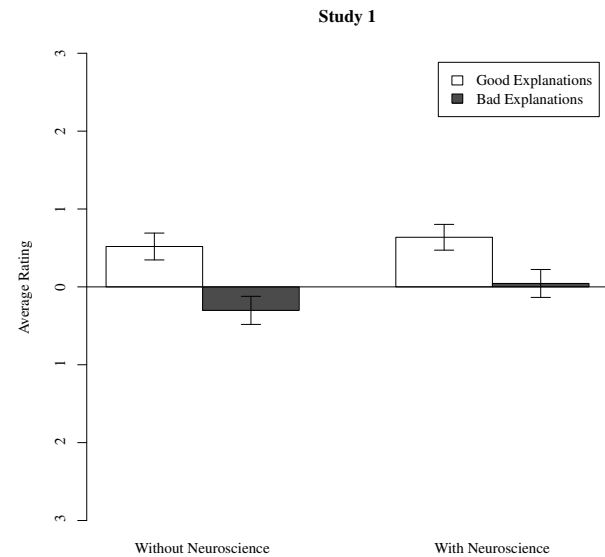
level of education (for the MTurk subjects) or class year and major (for the undergraduates).

2.2 Results

Unlike in Weisberg et al. (2008), some subjects in the current study received unequal numbers of good and bad explanations. In order to deal with this, we conducted a mixed-effects linear regression. The model included random intercepts by subject as well as random slopes by subject for the effect of Quality (the only within-subjects variable). We tested effects of Item, Group (MTurk or undergraduates), Length (long or short), Neuroscience (present or absent), and Quality (good or bad) and their interactions²; the model that best fit the data is shown in Table 1.

²Preliminary analyses revealed one effect of gender: an interaction between gender and explanation length. Men's ratings did not differ for long explanations ($M = 0.27, SD = 1.75$) and short explanations ($M = 0.33, SD = 1.74$). However, women rated long explanations ($M = 0.38, SD = 1.76$) more highly than short explanations ($M = -0.03, SD = 1.81$). We have no explanation for this unexpected gender difference, and because gender did not affect the other variables, we did not consider gender for the remainder

Figure 1: Average ratings of explanation quality in Study 1. Error bars represent 95% confidence intervals around the means.



This test revealed a main effect of Quality (Figure 1), where good explanations ($M = 0.58, SD = 1.68$) were rated more highly than bad explanations ($M = -0.13, SD = 1.80$). There was also a main effect of Neuroscience: Explanations that contained neuroscience ($M = 0.34, SD = 1.75$) were rated more highly than explanations that did not ($M = 0.11, SD = 1.80$). We also found a main effect of Length: Long explanations ($M = 0.34, SD = 1.76$) were rated more highly than short explanations ($M = 0.12, SD = 1.79$). Finally, there was a main effect of Group: MTurk workers ($M = 0.37, SD = 1.73$) gave overall higher ratings than undergraduates ($M = 0.10, SD = 1.80$).

The effects of Group, Neuroscience, and Quality also varied by item, as indicated by the significant interactions. To examine these interactions, we conducted separate linear regressions for each item examining main effects of Group, Neuroscience, Length, and Quality. The results are summarized in Table 2. Although the magnitudes (and therefore significance levels) of the effects varied by item, only two effects were not in the predicted directions; for Item 4, there were non-significant negative effects of Group and Neuroscience. The effects of neuroscience for Item 2 and quality for Item 4 were small and non-significant, but in the predicted directions.

of our analyses.

Table 2: Regression coefficients for individual item analysis in Study 1 (* $p < .05$, + $p < .10$).

	Item 1	Item 2	Item 3	Item 4
Length	0.04	0.05	0.17 ⁺	0.19*
Neuroscience	0.22*	0.08	0.34*	-0.10
Group	0.14	0.18*	0.30*	-0.06
Quality	0.69*	0.44*	0.23*	0.08

2.3 Discussion

Study 1 was designed to replicate the seductive allure effect and test for the contribution of explanation length. Subjects did indeed judge longer explanations as better than shorter ones overall, demonstrating a general bias towards longer explanations. However, this length preference does not fully explain the seductive allure of neuroscience. Making explanations longer does make them seem better, but adding neuroscience information does as well, and these two modifications had independent effects. This result confirms other recent studies that show that the seductive allure effect obtains when explanation length is controlled (Fernandez-Duque et al., 2015; Rhodes et al., 2014).

There was also a strong effect of explanation quality: Good explanations were judged as better than bad explanations overall. This result demonstrates that people are not generally confused about what makes certain explanations better than others and are able to distinguish between good and bad explanations. However, as noted above, the addition of neuroscience information interferes with this ability.

Finally, we found that undergraduates gave overall lower ratings than MTurk workers. This is likely not due to the undergraduates having a higher level of education, since 99% of the MTurk workers reported having at least some college education, and 50% reported earning an advanced degree. Instead, the experience of participating in research as part of a class, or of currently being a member of an educational community, may serve to increase overall skepticism. Regardless, both populations showed the same general pattern of responses to the explanations (i.e., there were no significant interactions with subject group).

These main effects appeared in nearly the same way for all four items; however, there were some differences in how subjects responded to the four phenomena. Notably, the phenomenon describing the differences between seeing and imagining objects (Item 4) did not show an effect of Quality or Neuroscience. This item was rated higher overall than the others (as indicated by the significant main effect for Item 4 in the regression), and there was little difference in ratings between the different versions of the explanation. In addition, the phenomenon describing attentional blink (Item 2) did not show as strong of an effect of Neuroscience. This

may be due to the fact that the neuroscience information in the explanations of this phenomenon is entirely contained in the first sentence, separate from the explanatory (or circular) information in the second sentence. This structure may have made it easier for subjects to see that the neuroscience information was not relevant to the explanation's quality.

Overall, Study 1 demonstrates that the seductive allure effect replicates and is not solely due to length. Study 2 begins to more directly address why the effect happens. To do so, rather than asking subjects to rate single explanations for a phenomenon, we ask them to choose which of two explanations they find most satisfying. Each pair contained a good explanation and a bad explanation, but either both contained neuroscience, neither contained neuroscience, or only the bad explanation contained neuroscience. This is a somewhat less ecologically valid design, since it is rare that people would need to evaluate multiple explanations for a single phenomenon. However, this design allows us to test directly how neuroscience information may interfere with people's ability to distinguish good from bad explanations. Given previous results, we expected that subjects would generally be able to distinguish good from bad explanations if both or neither contained neuroscience. If, however, neuroscience information has the effect of masking the poor quality of the bad explanations, people should be less likely to distinguish good from bad explanations when only the bad one contains neuroscience.

3 Study 2

3.1 Method

Subjects. This study included 130 undergraduates (86 female, 44 male; mean age = 19.5 years, range = 18–27) and 130 MTurk workers (90 female, 37 male, three unreported; mean age = 40.6 years, range = 19–71). The undergraduates were recruited from the psychology subject pool at the University of Pennsylvania and received course credit for their participation. The MTurk workers were recruited from Amazon's system and were paid 20 cents for their participation. An additional seven subjects (three MTurk workers and four undergraduates) were recruited but excluded from the final analyses for failing an attention check (described below).

Design. There were three between-subjects conditions in this study. As in Study 1, there were four trials per subject, each of which used a different phenomenon (order randomized). Each phenomenon was accompanied by both a good and a bad explanation. In the Without Neuroscience condition (41 MTurk workers and 43 undergraduates), neither explanation contained any neuroscience information, and in the With Neuroscience condition (42 MTurk workers and 45

undergraduates), both explanations contained neuroscience information. The crucial condition was the Mixed condition (47 MTurk workers and 42 undergraduates), in which the good explanation did not contain neuroscience information and the bad one did, pitting quality and neuroscience against each other.

Materials. We used the same four phenomena as in Study 1, accompanied by the short versions of the four possible explanations for each phenomenon: good and bad explanations both with and without irrelevant neuroscience information (see supplemental materials).

Procedure. Subjects completed an online survey distributed on Qualtrics. For each trial, they first read a description of a psychological phenomenon, which appeared in isolation on the screen for 10 seconds before they were allowed to advance to the next screen. On the second screen of each trial, the phenomenon appeared again at the top, followed by the prompt, “Please choose which explanation you find more satisfying.” Subjects always saw one good explanation and one bad explanation as well as the choice “both are equal.” The “equal” option always appeared in the center, with the left/right position of the good and bad explanations randomized across trials. After making their choice, subjects were asked to explain why they had made that choice in one or two sentences. There were four such trials in the experiment, each involving a different phenomenon and its accompanying explanations.

After the second of the four trials, subjects engaged in an attention check. Following methods recommended in Openheimer, Meyvis & Davidenko (2009), this checking trial presented another phenomenon and two explanations so that it looked superficially like the other four trials. For this trial, we included instructions at the end of the phenomenon description telling subjects to choose the “equal” option and to write that they had done so as their justification. As noted above, we excluded seven subjects who failed this check by selecting a different option. At the end of the survey, subjects provided the same demographic information as in Study 1: age, gender, and education level.

3.2 Results

To analyze the data, we conducted a mixed-effects logistic regression predicting whether subjects selected the good explanation on each trial, considering this response as correct and the other two responses (selecting the bad explanation or the “both are equal” option) as incorrect. The model included random intercepts by subject. We tested effects of Item, Group (MTurk workers or undergraduates), Condition (With Neuroscience, Without Neuroscience or Mixed), and possible interactions. Preliminary analyses found no effect

Table 3: Study 2 mixed-effects logistic regression model ($p < .05$).

Predictor	Estimate	SE	z
Intercept	0.05	0.14	0.33
Item 2	0.67	0.13	5.05*
Item 3	-0.25	0.12	-2.08*
Item 4	-0.94	0.13	-7.44*
Group	0.25	0.14	1.75
Without Neuroscience condition	0.63	0.21	3.08*
With Neuroscience condition	1.12	0.22	5.18*
Group x Item 2	-0.20	0.13	-1.52
Group x Item 3	0.19	0.12	1.53
Group x Item 4	0.37	0.12	3.05*
Without Neuroscience cond. x Group	-0.49	0.21	-2.37*
With Neuroscience cond. x Group	-0.53	0.21	-2.48*

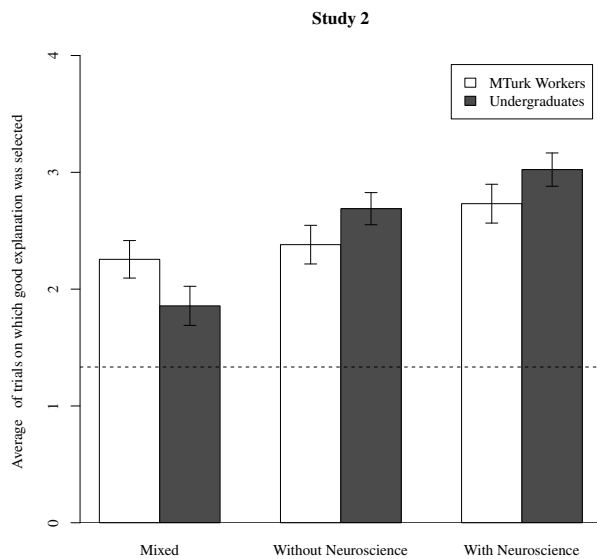
Note. The intercept represents the Mixed condition and undergraduate subjects. Item is deviation coded, such that the coefficient for each level represents deviation from the grand mean; Item 1 is the reference level.

of gender, so this variable is not considered further. The model that best fit the data is presented in Table 3.

The primary result from these analyses is the significant main effect of condition: Subjects found it more difficult to determine which was the good explanation in the Mixed condition where only the bad explanation contained neuroscience (Figure 2). Specifically, subjects were significantly more likely to select the good explanation in either the With Neuroscience (72.0% of trials) or Without Neuroscience (63.5% of trials) conditions than in the Mixed condition (51.7% of trials).

There was also a significant Group x Condition interaction. Follow-up regressions conducted on each group separately showed that the MTurk workers were significantly more likely to select the good explanation in the With Neuroscience condition as compared to the Mixed condition ($\beta = 0.58, p < .05$), but there was no significant difference between the Without Neuroscience and Mixed conditions ($\beta = 0.15, p = .59$). For undergraduate subjects, subjects selected the good option significantly more often in both the With Neuroscience ($\beta = 1.68, p < .001$) and Without Neuroscience ($\beta = 1.14, p < .001$) conditions than in the Mixed condition. Thus, although the main effect of condition was significant in the whole sample, it was driven more by the undergraduate subjects than by the MTurk workers. Finally, although there were significant differences between items, as in Study 1, there was not a significant Item x Condition or Item x Condition x Group interaction. Therefore, the effect

Figure 2: Average number of trials on which subjects selected the good explanation in Study 2. Error bars represent 95% confidence intervals around the means. The dotted line represents chance performance since selecting the good explanation was one of three possible responses on each trial.



of Condition and the Condition x Group interaction were not significantly different across items.

To gain further insight into subjects' responses, for the two conditions that used explanations containing neuroscience language (With Neuroscience and Mixed), we performed a text search on subjects' justifications for words related to neural processes generally ("brain", "lobe," "scan", "neur*") and for the specific neuroscience terms used in the explanations themselves ("premotor," "cortex"); 58% of justifications in these two conditions contained at least one of these terms. A research assistant, who was blind to condition, group, and study hypotheses, further coded these justifications for whether subjects referred to the brain as adding value to an explanation. For example, "this gives a biological explanation and uses brain parts to explain", "brain scans and timing seem more accurate", and "it's more in depth and seems more factual because it is talking about brain parts and stuff." Of justifications that referenced the brain, 84% did so in this positive way. The remaining justifications suggested a good grasp of the irrelevance of this information, such as these undergraduates: "I do not want to hear about what the brain is doing. I am interested in WHY the phenomenon occurs in a more general sense" and "Saying because of frontal lobe areas is not a sufficient explanation, but just states where the processing is occurring." Or, as one MTurk worker put it, "Talking mumbo jumbo about the

frontal lobes without explaining what is actually happening is bullshit."

Each subject was given a score (out of 4) for the number of positive brain-based justifications they gave. A 2 (Group: MTurk workers, undergraduates) x 2 (Condition: With Neuroscience, Mixed) ANOVA revealed only a significant main effect of Group, $F(1,166) = 7.08$, $p < .01$, $\eta^2 = .04$: Undergraduates ($M = 0.62$, $SD = 0.84$) were overall more likely than MTurk workers ($M = 0.31$, $SD = 0.66$) to refer to the brain as adding value to an explanation. There was neither a significant effect of Condition nor a significant Group x Condition interaction.

3.3 Discussion

When asked to choose between good and bad explanations of a psychological phenomenon, subjects selected the good explanation as being more satisfying on the majority of trials. These results confirm subjects' ratings from Study 1, in which subjects tended to rate the good explanations more positively. Taken together, these results demonstrate that subjects understand the difference between good and bad explanations.

The one exception to this conclusion is the Mixed condition, in which the bad explanation contained irrelevant neuroscience information and the good explanation did not. Here, subjects were seduced by the presence of neuroscience information, which made them less likely to prefer the good explanations than in the other conditions. Although the main effect of condition was significant in the full sample, it was driven primarily by the undergraduates. Indeed, the undergraduates' justifications were more likely to mention neuroscience in a positive light. This suggests that the presence of neuroscience information played a key role in convincing these subjects that the bad explanations were satisfactory. As the justifications quoted above illustrate, some undergraduates appeared to rely on the presence of neuroscience as a heuristic to judge the quality of an explanation.

It is not entirely clear why undergraduates would be more attracted to explanations containing neuroscience information than MTurk workers, although currently learning about psychological and neuroscientific phenomena might have swayed the undergraduates to lend more weight to the presence of neuroscience. One possibility is that, since they currently are learning about the functions of the brain, reading about specific phenomena in which the brain appears to play a causal role leads them to judge explanations with neuroscience explanations more favorably. In support of this argument, students who were currently taking a course on neuroscience (Weisberg et al., 2008, Study 2) showed a stronger attraction to explanations containing irrelevant neuroscience information than students recruited from the introductory psychology pool (Weisberg et al., 2008, Study 1).

Finally, subjects' performance in this study can start to explain one of the item effects seen in Study 1. Specifically, in Study 1, the seeing/imagining item (Item 4) was judged similarly regardless of quality or presence of neuroscience. Here in Study 2, about twice as many subjects chose the "both are equal" option for this item than for the other three, indicating that they could not see a difference between the good and bad explanations for this item. The difference between the two versions was very slight, changing only "uses the same process" to "results in the same array of responses". In fact, a number of participants explicitly stated in their justifications that the two explanations seemed the same (e.g., "Both explanations sound like they are saying the same thing", "Both explanations are similar and say the same thing just in a slightly different way.")

4 Study 3

Having determined that length does not underlie the seductive allure effect, and that neuroscience information is effective at disguising bad explanations, Study 3 tested another possible reason that neuroscience explanations are appealing to subjects, namely that this information tends to include technical jargon. If subjects are attracted to scientific-sounding terms, then neuroscience information *per se* is not seductive; subjects' responses can be influenced by the presence of any jargon. However, technical language and references to the brain were confounded in the explanations used thus far, preventing us from determining whether any reference to the brain would be sufficient or whether technical jargon is necessary. Study 3 constructed alternative versions of these explanations in order to tease out which type of information is responsible for the seductive allure effect.

This study's design mirrored that of Study 1, in which subjects read descriptions of psychological phenomena one at a time and then rated one explanation of each phenomenon. In Study 3, these explanations came from one of two sets: Simple Neuroscience, in which the explanations referred to brain scans and neural processes but in simple language without reference to specific brain areas, and Neuroscience Plus Jargon, in which the explanations included technical terms to refer to the type of brain scan used and the individual areas of the brain. In both cases, these stimuli were constructed by modifying the Short versions of the stimuli used in Study 1, and the Short-Without Neuroscience condition from that study serves as a control condition here.

This set of stimuli allows us to test among three hypotheses. If neuroscience information alone is responsible for the seductive allure effect, we should expect similar ratings for the Neuroscience Plus Jargon and the Simple Neuroscience explanations, both of which should be rated more highly than the Without Neuroscience explanations. If neuroscience information appeals because it contains fancy jar-

gon, then the Neuroscience Plus Jargon explanations should be rated more highly than the other two, which should not differ. Finally, there might be an additive effect of jargon and neuroscience language, in which case the Neuroscience Plus Jargon explanations would be rated more highly than the Simple Neuroscience explanations, which would in turn be rated more highly than the Without Neuroscience explanations.

4.1 Method

Subjects. The final sample for this study included 88 undergraduates (63 female, 25 male; mean age = 19.5 years, range = 18–22) and 82 MTurk workers (42 female, 38 male, two unreported; mean age = 35.0 years, range = 19–67). As in previous studies, the undergraduates were recruited from the psychology subject pool at the University of Pennsylvania and received course credit for their participation. The MTurk workers were recruited from Amazon's system and were paid 20 cents for their participation. An additional 50 subjects completed the survey but were excluded from the final analyses for failing an attention check (described below; 22 MTurk workers and 28 undergraduates). Although more subjects failed the attention check here than in Study 2, the design of this study was different and may have presented a less engaging task than Study 2, and these numbers are in line with other studies that included similar attention checks (see Oppenheimer et al., 2009).

Design. This study used a 2 (Group: MTurk, undergraduate) \times 2 (Neuroscience: Simple Neuroscience, Neuroscience Plus Jargon) \times 2 (Quality: good, bad) design. Group and Neuroscience were between-subjects variables and Quality was a within-subjects variable.³ Subjects were assigned to either the Neuroscience Plus Jargon condition (42 MTurk workers and 44 undergraduates) or the Simple Neuroscience condition (40 MTurk workers and 44 undergraduates). Data from the 45 MTurk workers and 45 undergraduates in the Without Neuroscience-Short condition from Study 1 were also used here for comparison.

Materials. To construct the stimuli, we used the same four psychological phenomena as in Studies 1 and 2, and modified the With Neuroscience-Short explanations to fit the new conditions (see supplemental materials). Explanations in the Simple Neuroscience condition removed references to specific brain-scanning techniques or brain areas and replaced them with generic terms ("brain scans", "visual area"). Explanations in the Neuroscience Plus Jargon condition enhanced existing references to include as much specific jar-

³As in Study 1, the randomization algorithm led to 11 subjects receiving either good explanations on every trial or bad explanations on every trial. The inclusion of these subjects did not affect any analyses, so they were left in the sample.

gon as possible (“fMRI scans”, “parietal lobe”). Each explanation had a good and a bad version, and this modified information was exactly the same in both versions.

Procedure. All subjects filled out an online survey on Qualtrics. As in Study 1, for each of the four trials, subjects first read a description of one of the four psychological phenomena, which appeared on the screen for 10 seconds before they were allowed to advance. On the second screen, this phenomenon description appeared again, followed by one of the four possible explanations of the phenomenon, according to the subject’s assigned condition; whether they saw the good or bad version of the explanation was randomly determined on each trial (as described above in the Design section). Subjects were asked to rate how satisfying they found the explanation on a -3 (very unsatisfying) to $+3$ (very satisfying) scale. They were then asked to justify their rating in one or two sentences.

As in Study 2, after the first two trials, we included an attention check. This attention check presented another description of a psychological phenomenon and an explanation for it in exactly the same way as the other trials, except that the last sentence of the explanation told subjects to select 3 on the scale. As noted above, 50 subjects failed this attention check (by failing to select 3 and/or by demonstrating a lack of attentiveness in their justifications for this item) and are not included in our analyses. At the end of the survey, subjects responded to the same basic demographic questions as in Studies 1 and 2, reporting their age, gender, and highest level of education.

4.2 Results

In order to have a control condition with which to compare the current subjects’ responses, the analyses for this study additionally include the responses from the subjects in the Without Neuroscience-Short condition from Study 1 (45 MTurk workers and 45 undergraduates). Preliminary analyses revealed no effects of gender, so it was not included in our analyses.

As in Study 1, we conducted a mixed-effects linear regression analysis. The model included random intercepts by subject as well as random slopes by subject for the effect of Quality (the only within-subjects variable). We created two dummy variables to examine the effects of neuroscience and jargon; the Neuroscience variable coded whether neuroscience information was present (the Simple Neuroscience and Neuroscience Plus Jargon conditions) or absent (the Without Neuroscience condition). Similarly, the Jargon variable coded whether jargon was present (the Neuroscience Plus Jargon condition) or absent (the Simple Neuroscience and Without Neuroscience conditions). The regression tested effects of Item, Group (MTurk or undergraduates), Neuroscience (present or absent), Jargon (present or

Table 4: Study 3 mixed-effects linear regression model (* $p < .05$).

Predictor	Estimate	[95% CI]	t
Intercept	0.02	[-0.19, 0.24]	0.17
Item 2	0.33	[0.06, 0.62]	2.19*
Item 3	-0.91	[-1.24, -0.62]	-5.97*
Item 4	0.98	[0.69, 1.25]	6.44*
Neuroscience	0.27	[0.05, 0.51]	2.44*
Group	0.15	[0.05, 0.27]	2.64*
Quality	0.57	[0.35, 0.78]	5.39*
Neuroscience x Item 2	-0.37	[-0.68, -0.07]	-2.42*
Neuroscience x Item 3	-0.04	[-0.36, 0.27]	-0.23
Neuroscience x Item 4	-0.05	[-0.38, 0.28]	-0.35
Quality x Item 2	0.12	[-0.18, 0.44]	0.82
Quality x Item 3	-0.39	[-0.67, -0.07]	-2.62*
Quality x Item 4	-0.57	[-0.85, -0.25]	-3.83*

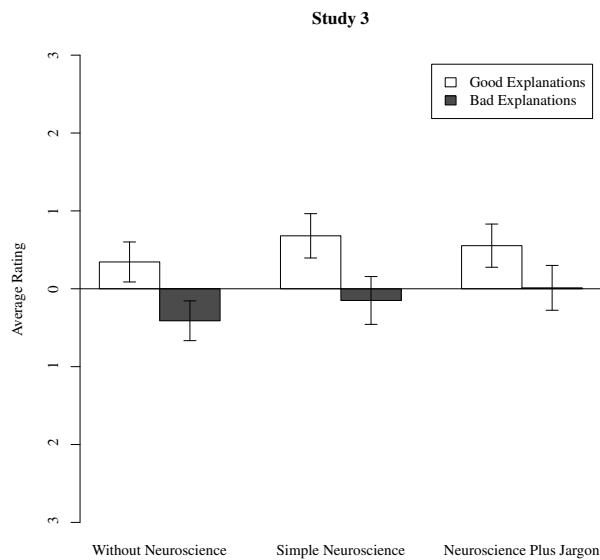
Note: This regression predicted subjects’ ratings of the quality of the explanations. The intercept represents the explanations that did not contain neuroscience, undergraduate subjects, and bad explanations. Item is deviation coded, such that the coefficient for each level represents deviation from the grand mean; Item 1 is the reference level.

absent), and Quality (good or bad) and their interactions; the model that best fit the data is shown in Table 4.

The analysis revealed significant main effects of Group, Quality, and Neuroscience (Figure 3). MTurk subjects ($M = 0.33$, $SD = 1.90$) gave higher overall ratings than undergraduate subjects ($M = 0.03$, $SD = 1.89$), and subjects rated good explanations ($M = 0.52$, $SD = 1.86$) more highly than bad explanations ($M = -0.19$, $SD = 1.88$), replicating the results of Study 1. The significant main effect of Neuroscience indicates that explanations were rated more highly in the two conditions that used neuroscience language ($M = 0.29$, $SD = 1.95$) than in the Without Neuroscience condition ($M = -0.03$, $SD = 1.79$). There was no significant effect of Jargon, meaning that the Neuroscience Plus Jargon condition ($M = 0.29$, $SD = 1.91$) was not significantly different from the other two combined ($M = 0.12$, $SD = 1.89$).

As in Study 1, the effects of Neuroscience and Quality also varied by item, as indicated by significant interactions. To examine these interactions, we conducted separate linear regressions for each item examining main effects of Group, Neuroscience, Length, and Quality. The results are summarized in Table 5. Although the magnitudes (and therefore significance levels) of the effects varied by item, only two effects were not in the predicted directions; consistent with Study 1, Item 2 had a negative, non-significant effect

Figure 3: Average ratings of explanation quality in Study 3, including the Without Neuroscience condition from Study 1. Error bars represent 95% confidence intervals around the means.



of Neuroscience, and Item 4 had a negative, non-significant effect of Quality.

To analyze subjects' justifications, we searched for all references to neuroscience, as in Study 2; 24% of all justifications referenced the brain, and 58% of those did so in a positive manner. A 2 (Group: MTurk workers, undergraduates) x 2 (Condition: Simple Neuroscience, Neuroscience Plus Jargon) ANOVA revealed no significant effects on subjects' average number of positive brain-based justifications.

However, subjects' justifications provide further insight into one of the item effects. Many of the justifications for the attentional blink item (Item 2), which did not show a significant neuroscience effect, mentioned that the neuroscience information seemed unconnected with the phenomenon: e.g., "The explanation mentions the frontal lobe areas but does not really say how the areas relate to attentional blink", and "The explanation does not explain how frontal areas are related to the temporal relationship between the two houses." This item effect was consistent with the findings from Study 1, and justifications such as this support our suggestion that putting the information about the frontal lobe in a separate sentence may have made it easier for subjects to separate this information from the body of the explanation, explaining the lesser effect of neuroscience for this item.

Table 5: Regression coefficients for individual item analysis in Study 3 (* $p < .05$).

	Item 1	Item 2	Item 3	Item 4
Neuroscience	0.27*	-0.09	0.24*	0.22*
Quality	0.55*	0.69*	0.17	-0.01
Group	0.04	0.32*	0.18	0.08

4.3 Discussion

Explanations with neuroscience information, whether presented as simply as possible without jargon or with reference to specific neural techniques and brain areas, were more satisfying than explanations without neuroscience information. However, there was no difference between the two neuroscience conditions. This suggests that any reference to neuroscience is sufficient to cause the effect, and that adding technical jargon does not increase subjects' ratings.

Additionally, subjects judged good explanations more highly than bad ones, and undergraduate subjects gave overall lower ratings than MTurk workers. These effects replicate the findings of Study 1 and suggest two additional conclusions. First, people can generally discriminate good from bad explanations. Second, participating in research as part of one's educational experience seems to make subjects more skeptical overall, but does not eliminate the seductive allure effect.

These main effects were significant overall, but varied somewhat by item. As in Study 1, the attentional blink item (Item 2) did not show an effect of neuroscience information and the seeing/imagining item (Item 4) did not show an effect of quality. In the case of the former, as noted above, the neuroscience information was contained in a separate sentence rather than being directly linked to the explanatory information. This may have made it easier for subjects to realize that this information was irrelevant. In the case of latter, as in Studies 1 and 2, subjects seemed generally unable to tell the difference between the good and bad versions of this item.

5 General discussion

The seductive allure effect of neuroscience, first observed by Weisberg et al. (2008), occurs when subjects judge that explanations for psychological phenomena (especially bad ones) that contain irrelevant neuroscience information are better than explanations that do not. The current studies provide new insight into why this effect happens.

First, although the original stimuli used to demonstrate this effect confounded neuroscience information with length, our Study 1 and independent replications by

Fernandez-Duque et al. (2015) and Rhodes et al. (2014; 2015) using different methods show that length does not account for the effect. Subjects do judge longer explanations as significantly better, but they also judge explanations with neuroscience information as significantly better when length has been controlled for. Something about neuroscience information itself, then, is responsible for the effect.

Study 2 showed that, although subjects generally chose correctly when explicitly comparing good and bad explanations, subjects were still seduced into choosing the bad explanation when it contained neuroscience information but the good explanation did not. Surprisingly, undergraduates' justifications indicate that they explicitly used the presence of neuroscience as a marker of a good explanation. These results thus provide an especially direct demonstration of the power of neuroscience information for this population. These results also suggest that education in the field of psychology, at least at the introductory level, might aggravate the effect. Further, results from Study 3 suggest that it is not simply fancy terms or scientific jargon that seduces subjects. Rather, any reference to the brain was sufficient to make an explanation seem more satisfying than a logically parallel explanation without any such references.

Having eliminated these potential explanations for the seductive allure effect, we are left with the general questions of why this effect happens and of whether it is unique to psychology. One possibility is that people are generally skeptical about psychology (Ferguson, 2015; Keil, Lockhart & Schlegel, 2010; Lilienfeld, 2012), believing that its investigative methods do not justify it as a "real science". An extreme version of this skepticism would endorse explanations that eliminate psychological terms altogether and utilize only neuroscience vocabulary. According to this theory, explanations that reference "harder" sciences may be seen as generally better across disciplines, but will have a more pronounced effect in psychology because of a general bias towards making psychological explanations sound "more scientific". Indeed, this may have been the strategy adopted by the undergraduates in Study 2, who explicitly reported liking explanations more when they contained brain-based language.

A second possibility is that people are intuitively dualist. Even though people may explicitly assert that the brain is involved in cognitive tasks, rejecting a strict Cartesian substance dualism, they may nevertheless fail to acknowledge the causal role of the brain in all aspects of our mental and emotional lives (Bloom, 2004, 2006). For example, in a recent study, subjects were told about a hypothetical machine that could perfectly duplicate people or animals, down to the very last cell (Forstmann & Burgmer, 2015). Subjects typically said that physical traits, such as scars or illnesses, would be preserved in the copy, but they were less likely to say the same for mental traits, such as emotions or memories. Results like this suggest that people believe that some-

thing over and above the physical brain is at least partly responsible for thoughts and feelings. They may thus be attracted to neuroscience information because they find it surprising and compelling when neural activity is shown to underlie mental activity. On this view, the seductive allure effect may be unique to psychology, since issues of dualism do not generally arise in other sciences.

A third possibility is that people may see explanations that contain neuroscience as providing additional causal information. Previous work has shown that people are sensitive to descriptions of causes. For example, subjects are less likely to ignore base rates when provided with information that causally links the base rates to the target outcome (Tversky & Kahneman, 1982). People are also particularly biased towards teleological information, which provides evidence of an ultimate cause for an event (Kelemen, 1999; Lombrozo & Carey, 2006; Lombrozo, Kelemen & Zaitchik, 2007). Differential responses to our four stimulus items also suggest that stronger effects obtain when the neuroscience information is described as causally related to the explanation (e.g., "the parietal lobe *governed* the babies' expectations ..."). If brain processes are seen as providing an underlying cause for the psychological phenomena in question, then their appeal may be due to this general bias. Unlike the previous two possibilities, if this explanation for the effect is correct, then the effect should appear across a range of sciences.

Finally, it is possible that neuroscience seduces because of a general preference for reductive explanations (Craver, 2007; Garfinkel, 1981; Trout, 2007). These explanations recruit the vocabularies and methods of scientific disciplines that are considered more fundamental (Oppenheim & Putnam, 1958). This is not in itself an error; explanations with a reductionist form are often of high quality. But the explanations in this case have this form without any accompanying content, since the reductionist (neuroscience) information did not provide any additional explanatory power.

If reductionism is indeed the key to explaining the seductive allure effect, then neuroscience information produces the effect because people see psychology as dependent upon neuroscience to verify its claims. If this is the case, then this effect should not be unique to psychological phenomena, but rather should appear across a variety of sciences. For example, an explanation of a biological phenomenon might be seen as more satisfying when it includes references to chemical processes, even if the chemical information is irrelevant. An ongoing study is investigating this hypothesis, drawing phenomena and explanations from across the social and natural sciences to test whether the seductive allure effect may appear in different disciplines. Results from this study can cast additional light on why the effect happens.

Insights from experts can also help with this effort: Neuroscience experts in Weisberg et al. (2008) were not seduced by information that came from their own domain of exper-

tise (see also Eriksson, 2012), suggesting that increased education can be an effective antidote. If experts are immune to some aspects of the seductive allure effect, their responses can provide insight into how to prevent it.

Regardless of whether the seductive allure effect is specific to psychology or also appears in other fields, it has important implications for how scientific information is communicated to the public (Weisberg, 2008). Since some individuals may use the presence of neuroscience information as a marker of a good explanation, like the undergraduates who participated in Study 2, it is imperative to find ways to increase general awareness of the proper role for neuroscience information in explanations of psychological phenomena. The present studies suggest that this effect is robust against changes to an explanation's length and to the terms in which the neuroscience information is described, implying that preventing the seductive allure effect from happening may be difficult. Future studies should continue to investigate why neuroscience information is so alluring, and to what types of subjects, in order to combat the superfluous appeals to neuroscience that are currently popular—and convincing—in public debates.

6 References

- Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews Neuroscience*, *11*, 284–292. <http://dx.doi.org/10.1038/nrn2795>.
- Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.
- Bloom, P. (2006). Seduced by the flickering lights of the brain. *Seed Magazine*. Retrieved from http://seedmagazine.com/content/article/seduced_by_the_flickering_lights_of_the_brain/.
- Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, *26*(8), 4–16. <http://dx.doi.org/10.3102/0013189X026008004>.
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*, 9–64. <http://dx.doi.org/10.1257/0022051053737843>.
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.
- Eriksson, K. (2012). The nonsense math effect. *Judgment and Decision Making*, *7*(6), 746–749.
- Farah, M. J. (2012). Neuroethics: The ethical, legal, and societal impact of neuroscience. *Annual Review of Psychology*. <http://dx.doi.org/10.1146/annurev.psych.093008.100438>.
- Farah, M. J., & Hook, C. J. (2013). The seductive allure of “seductive allure.” *Perspectives on Psychological Science*, *8*(1), 88–90. <http://dx.doi.org/10.1177/1745691612469035>.
- Ferguson, C. J. (2015). “Everybody knows psychology is not a real science”: Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist*, *70*(6), 527–542. <http://dx.doi.org/10.1037/a0039405>.
- Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S. D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*, *27*(5), 926–944. http://dx.doi.org/10.1162/jocn_a_00750.
- Forstmann, M., & Burgmer, P. (2015). Adults are intuitive mind-body dualists. *Journal of Experimental Psychology: General*, *144*(1), 222–235. <http://dx.doi.org/10.1037/xge0000045>.
- Garfinkel, A. (1981). *Forms of explanation*. New Haven, CT: Yale University Press.
- Goswami, U. (2006). Neuroscience and education: From research to practice? *Nature Reviews Neuroscience*, *7*(5), 406–411. <http://dx.doi.org/10.1038/nrn1907>.
- Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*, 1775–1785. <http://dx.doi.org/10.1098/rstb.2004.1546>.
- Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of Science*, *21*(8), 938–948. <http://dx.doi.org/10.1177/0963662512454072>.
- Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience*, *25*, 1397–1405. http://dx.doi.org/10.1162/jocn_a_00407.
- Keehner, M., Mayberry, L., & Fischer, M. H. (2011). Different clues from different views: The role of image format in public perceptions of neuroimaging results. *Psychonomic Bulletin & Review*, *18*(2), 422–428. <http://dx.doi.org/10.3758/s13423-010-0048-7>.
- Keil, F. C., Lockhart, K. L., & Schlegel, E. (2010). A bump on a bump? Emerging intuitions concerning the relative difficulty of the sciences. *Journal of Experimental Psychology: General*, *139*(1), 1–15. <http://dx.doi.org/10.1037/a0018319>.
- Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, *35*(6), 1440–1452. <http://dx.doi.org/10.1037/0012-1649.35.6.1440>.
- Kikas, E. (2003). University students' conceptions of different physical phenomena. *Journal of Adult Devel-*

- opment, 10(3), 139–150. <http://dx.doi.org/10.1023/A:1023410212892>.
- Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, 67(2), 111–129. <http://dx.doi.org/10.1037/a0023963>.
- Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99, 167–204. <http://dx.doi.org/10.1016/j.cognition.2004.12.009>.
- Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science*, 18(11), 999–1006. <http://dx.doi.org/10.1111/j.1467-9280.2007.02015.x>.
- McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, 107(1), 343–352. <http://dx.doi.org/10.1016/j.cognition.2007.07.017>.
- Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review*, 20(4), 720–725. <http://dx.doi.org/10.3758/s13423-013-0391-6>.
- Morse, S. J. (2004). New neuroscience, old problems. In B. Garland (Ed.), *Neuroscience and the law: Brain, mind and the scales of justice* (pp. 157–198). New York: Dana Press.
- Morse, S. J. (2011). The future of neuroscientific evidence. In *The future of evidence: How science and technology will change the practice of law* (pp. 137–164). Washington, DC: ABA Book Publishing.
- Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 2, pp. 3–36). Minneapolis, MN: University of Minnesota Press.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872. <http://dx.doi.org/10.1016/j.jesp.2009.03.009>.
- Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1432–1440. <http://dx.doi.org/10.1037/a0036844>.
- Rhodes, R. E., & Shah, P. (2015). Seeing behavior through the brain: Evidence of neurorealism, Manuscript under review.
- Rips, L. (2002). Circular reasoning. *Cognitive Science*, 26(6), 767–795. [http://dx.doi.org/10.1016/S0364-0213\(02\)00085-X](http://dx.doi.org/10.1016/S0364-0213(02)00085-X).
- Roskies, A. (2002). Neuroethics for the new millenium. *Neuron*, 35(1), 21–23. [http://dx.doi.org/10.1016/S0896-6273\(02\)00763-8](http://dx.doi.org/10.1016/S0896-6273(02)00763-8).
- Saks, M. J., Schweitzer, N. J., Aharoni, E., & Kiehl, K. A. (2014). The impact of neuroimages in the sentencing phase of capital trials. *Journal of Empirical Legal Studies*, 11(1), 105–131. <http://dx.doi.org/10.1111/jels.12036>.
- Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. New York: Basic Books.
- Schweitzer, N. J., Saks, M. J., Murphy, E. R., Roskies, A. L., Sinnott-Armstrong, W., & Gaudet, L. M. (2011). Neuroimages as evidence in a mens rea defense: No impact. *Psychology, Public Policy, and Law*, 17(3), 357–393. <http://dx.doi.org/10.1037/a0023581>.
- Scurich, N., & Shniderman, A. (2014). The selective allure of neuroscientific explanations. *PLOS One*, 9(9), e107529. <http://dx.doi.org/10.1371/journal.pone.0107529>.
- Sperber, D. (2010). The guru effect. *Review of Philosophy and Psychology*, 1(4), 583–592. <http://dx.doi.org/10.1007/s13164-010-0025-0>.
- Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, 69, 212–233.
- Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass*, 2, 564–591. <http://dx.doi.org/10.1111/j.1747-9991.2007.00081.x>.
- Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty* (pp. 153–160). New York: Cambridge University Press.
- Weisberg, D. S. (2008). Caveat lector: The presentation of neuroscience information in the popular media. *Scientific Review of Mental Health Practice*, 6(1), 51–56.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, 20(3), 470–477. <http://dx.doi.org/10.1162/jocn.2008.20040>.