When Is Science Considered Interesting and Important?

Samuel G. B. Johnson[1], Amanda Royka[2,4], Peter McNally[3] & Frank C. Keil[4]
(sgbjohnson@gmail.com, amanda.royka@yale.edu, peterkmcnally1@gmail.com, frank.keil@yale.edu)
[1] School of Management, University of Bath, Bath, BA2 7AY UK
[2] School of Chemical and Biological Sciences, Queen Mary University of London, London E1 4NS UK
[3] Social Norms Group, University of Pennsylvania, Philadelphia, PA 19104 USA
[4] Department of Psychology, Yale University, New Haven, CT 06520 USA

Abstract

Scientists seek to discover truths that are interesting and important. We characterized these notions by asking laypeople to assess the importance, interestingness, surprisingness, practical value, scientific impact, and comprehensibility of research reported in the journals *Science* and *Psychological Science*. These judgments were interrelated in both samples, with interest predicted by practical value, surprisingness, and comprehensibility, and importance predicted mainly by practical value. However, these judgments poorly tracked the academic impact of the research, measured by citation counts three and seven years later. These results suggest that although people have internally reliable notions of what makes science interesting and important, these notions do not track scientific findings' actual impact.

Keywords: Folk science; science methodology; interest; philosophy of science; scientometrics

Introduction

The scientific enterprise aims to uncover eternal truths, and psychological science seeks to understand the most fundamental aspects of the human condition. From our modern vantage point, we can see clearly which scientific theories and results have stood the test of time, as truly foundational scientific achievements—Euclid's explication of geometry, Newton's laws of motion, Smith's insights about economic activity, Darwin's theory of evolution are among the timeless truths that clarify the structure of the natural and social worlds. But as scientists in the trenches, it is much more difficult for us to know what research is truly significant. Mendel's insights into genetics were ignored in his day, revolutionizing biology only decades later. The full significance of Bayes' contributions to statistics long eluded the profession.

Given these difficulties, scientists are likely to develop heuristics to evaluate scientific importance (Kahneman & Frederick, 2002). One particularly plausible heuristic is the *counterintuitiveness* or *surprisingness* of the research finding. For example, research on cultural narratives finds that minimally counterintuitive myths (relying mainly on intuitive concepts, peppered with a few counterintuitive ones) are most likely to be remembered and passed on (Norenzayan et al., 2006). Indeed, at a very general level, learning is likeliest to occur when the difference between expectations and reality (i.e., prediction error) is largest (Rescorla & Wagner, 1972).

Surprise is often a good criterion for scientific importance. We might consider a scientific result to be important when it falsifies an element of a theory (Popper, 1959/1934) or requires us to reconceptualize a topic of inquiry altogether (Kuhn, 1962). In Bayesian terms, a result is highly diagnostic when it is highly improbable on most theories but highly probable on another (i.e., $P(E|H)$ is high but $P(E)$ is low). As heuristics go, the extent to which a result encourages theory change is an excellent proxy for scientific importance.

But often, this heuristic can lead us astray. This is because even practicing scientists have scientific theories and intuitive theories that co-exist in their minds (Goldberg & Thompson-Schill, 2009; Shtulman & Varcarcel, 2012). Thus, although the disagreement between a result and existing *scientific* theory is a plausible proxy for scientific importance, disagreement with one's *lay* theory is not, if it is superseded by one's scientific understanding. For example, suppose that a psychologist believes that our behavior is guided by the unconscious activation of stereotypes, as suggested in the social priming literature. These original effects are highly counterintuitive, and if true, of great scientific significance. However, even though conceptual replications of these priming effects (e.g., using different stereotypes) would no longer contradict *scientific theory* (assuming we accept the initial demonstration), they would remain counterintuitive relative to our *folk theory*. Thus, this creates a misalignment between the scientific and lay surprisingness of a particular finding. To the extent that scientists rely on their folk theories rather than their scientific understanding for evaluating whether a finding is surprising, they may share this misalignment.

Regardless of the normative importance of counterintuitiveness, there is no question that many scientists prize it highly. Scientists, particularly during training, are often advised to seek out counterintuitive results. For example, one guide to doing "interesting" research advises (Gray & Wegner, 2013; pg. 550):

> One concrete test for evaluating ideas is to imagine the most surprising outcome possible (i.e., the best case scenario). If results were exactly as predicted, would they be interesting? If not, you should dream bigger when hypothesizing or perhaps consider the opposite of your hypothesis—if one way is intuitive, the opposite may be surprising.

Whose intuitions are we trying to contradict? "Grandmothers, not scientists," note the authors: "Ideally, research should counter both scientists' and laypeople's intuitions, but we emphasize the latter" (pg. 550). It is

hard to disagree with this as career advice, but it nonetheless raises uncomfortable concerns about replicability. After all, results with low prior probability are less likely to be true. Indeed, surprisingness is among the factors most associated with failure to replicate (Open Science Collaboration, 2012). It is presumably for this reason that the submission guidelines for *Psychological Science* now distinguish explicitly between "theoretical significance" (which is an acceptance criterion) and "surprising novelty" (which is not).

In this paper, we test two sets of issues, with Study 1 examining the folk science surrounding psychological research and Study 2 examining the natural sciences.

First, we ask what factors drive laypeople's judgments of how interesting and important scientific findings are. The opinions of laypeople, while likely divergent from experts, are important for two reasons. One reason is that scientists are laypeople in all fields aside from their own, and even in their own field may have lay intuitions that conflict with their scientific understanding of the field (Goldberg & Thompson-Schill, 2009). Thus, lay intuitions can creep into scientists' evaluations of research. A second reason is that the opinions of laypeople directly impact what scientific research is conducted, since laypeople are the ultimate consumers of taxpayer-funded research and since many scientists prioritize newsworthiness (to laypeople) in choosing topics to investigate. We study, therefore, the relative importance of surprisingness, perceived scientific impact, perceived practical value, and comprehensibility in guiding judgments of importance and interest.

Second, we ask how well these judgments track the objective academic impact of scientific findings, as quantified by their citations. Is the advice quoted above—to prioritize counterintuitiveness to laypeople—sound, if one's goal is to generate citations? Gray and Wegner (2013) suggest that it may be counterintuitiveness to scientists that drives citations in the short term, but to laypeople that drives citations in the longer term. We begin to examine this issue by looking separately at citation counts 4- and 7-years post-publication, testing whether lay judgments predict such measures of impact.

Study 1

In our first study, we looked at the factors influencing judgments of interest and importance, as well as citation counts, for articles published in *Psychological Science*.

Method

Participants. We recruited 60 participants from Amazon Mechanical Turk. Across our two studies, 57% of participants were female, 42% had completed at least a 4-year college degree, and the average age was 35. Only 8% of participants had doctoral-level training in any field, so the vast majority of participants were laypeople in the specific fields featured in our studies.

Participants were excluded if they incorrectly answered

more than 30% of a set of 20 check questions ($N = 8$).

Materials. The materials were derived from abstracts of 40 articles appearing in the journal *Psychological Science* in the January, February, and March 2012 issues. A power analysis, treating item as the unit of analysis (like our main analysis below), revealed that 40 items is sufficient to detect correlations of $r > .41$ with 80% power.

For each abstract, a short summary was developed by the second author. For example, the actual abstract of one article (Frankenstein et al., 2012) read:

> We examined how a highly familiar environmental space—one's city of residence—is represented in memory. Twenty-six participants faced a photo-realistic virtual model of their hometown and completed a task in which they pointed to familiar target locations from various orientations. Each participant's performance was most accurate when he or she was facing north, and errors increased as participants' deviation from a north-facing orientation increased. Pointing errors and latencies were not related to the distance between participants' initial locations and the target locations. Our results are inconsistent with accounts of orientation-free memory and with theories assuming that the storage of spatial knowledge depends on local reference frames. Although participants recognized familiar local views in their initial locations, their strategy for pointing relied on a single, north-oriented reference frame that was likely acquired from maps rather than experience from daily exploration. Even though participants had spent significantly more time navigating the city than looking at maps, their pointing behavior seemed to rely on a north-oriented mental map.

We anticipated that real scientific abstracts like this one would be too long, syntactically complex, and jargon-filled to be comprehensible by most laypeople. Therefore, our summary version read:

> When presented with a virtual model of their hometown, people are able to more accurately point to familiar target locations when the people were oriented north and become progressively less accurate as they were oriented away from north. This suggests that people rely on a mental map that is oriented northward when trying to locate familiar places.

Comparable summaries were constructed for all 40 abstracts. Summaries were written at a minimum Flesch–Kincaid grade level of 12 and were of similar length.

We conducted pretests to ensure as strong of a perceived correspondence between the real abstract and summary as possible. In an initial pretest, each participant was assigned to read 10 of the abstracts along with their summaries, and rated their correspondence on a 0 ("A very poor match") to 10 ("An excellent match") scale. Any abstract with a score below 7 was targeted for revision and re-normed in a second pretest. All correspondences were rated above the scale midpoint in the second pretest (except one item which was omitted due to a coding error).

As an objective measure of academic impact, we obtained the Google Scholar citation counts for each article approximately 4 years (on 26 March 2016) and 7

years post-publication (on 29 January 2019) (on the pros and cons of Google Scholar versus other bibliometric databases, see Harzing & Alakangas, 2016). These were square root transformed, to account for the skewness of citation data.

Procedure. Participants each viewed 10 of the 40 summaries (balanced across participants). For each finding, participants first read the summary and then, on subsequent pages, made six ratings:

Interest. How interesting are these findings to you?

Importance. How important do you think these findings are?

Surprise. How surprising do you think these findings are?

Scientific impact. How much do you think these findings will change the way scientists think about this topic?

Practical value. How useful do you think this finding is on a practical level?

Comprehensibility. How well do you think that you understand the description of this finding?

These ratings were all made on scales from 0 ("Not at all…") to 10 ("Very…"). Each rating was made on a separate page, with the summary repeated at the top of each page. The order of the interest and importance questions was counterbalanced across participants, and the other ratings were always made in the order above.

After the main task, participants checked off, from a list of 20 concepts, those that had appeared in the summaries. Participants incorrectly answering more than 30% of these check questions were excluded to decrease noise due to inattentiveness.

Results

Overall, participants' judgments were internally reliable, with significant correlations among many of our measures. However, these scores had little external predictive power: Citations 4 and 7 years later were not predicted by any judgment except comprehensibility.

First-order correlations. We averaged, for each item, across participants' ratings, and used these item-level means for our analyses. The first-order Pearson correlations among all measures are summarized in Table 1. Before probing these associations more carefully using regression models, we make two observations.

First, judgments of importance and interest were highly correlated, $r(38) = .59$, $p < .001$. Since these results are observational, this is consistent with several causal orders. It could be that importance is the more fundamental judgment, and these appraisals feed into interest. This would be consistent with the fact that usefulness judgments were even more strongly associated with importance, $r = .79$, than with interest, $r = .61$. Alternatively, interest could be the more fundamental judgment, with importance less natural to judge and confabulated in line with personal interest. Finally, these two assessments could be relatively independent,

depending on a mix of the same factors (such as usefulness) and differentiating factors (such as comprehensibility, which is only associated with interest).

	In	Im	Su	SI	PV
Im	.60***	—			
Su	.59***	.49**	—		
SI	.55***	.59***	.76***	—	
PV	.61***	.79***	.36*	.56***	—
Co	.66***	.11	.16	−.05	.17

º < .10	* < .05	** < .01	*** < .001

Note. Entries are first-order correlations among interest (In), importance (Im), surprise (Su), scientific impact (SI), practical value (PV), and comprehensibility (Co).

Table 1: First-order correlations (Study 1).

Second, in preparation for modeling interest and importance, we note that some of the other variables are strongly correlated, which can lead to a multicollinearity problem. Variance Inflation Factors were acceptable (VIF < 1.5 for the Step 1 models in Tables 2 and 6) for models that did not simultaneously include both surprise and scientific impact, which were correlated very highly, $r = .76$. This very high correlation suggests, perhaps not itself surprisingly, that laypeople tend to substitute the difficult question of what evidence tends to change scientists' theories with the easier question of what they personally find surprising (Kahneman & Frederick, 2002). To address this problem, we omitted the scientific impact variable from the models. We included surprise rather than scientific impact since this seems to be the more natural assessment, but the results are similar if we instead include scientific impact or the average of the two.

Predictors of interest and importance. Table 2 shows the regression coefficients predicting judgments of interest. The Step 1 model uses surprise, practical value, and comprehensibility to model interest, and the Step 2 model adds importance to capture any added value.

As shown in the regression table, the strongest predictor of interest was comprehensibility, followed by practical value, followed by surprise, but all three predictors were highly significant, making independent contributions to interest. Together, these factors accounted for 80% of the variance in interest across items. Adding importance did not add any predictive power.

DV: Interest		
	Step 1	**Step 2**
Su	.26 (.06)***	.23 (.06)***
PV	.33 (.07)***	.23 (.10)*
Co	.42 (.06)***	.43 (.06)***
Im		.15 (.12)
R^2	.802	.811

DV: Importance		
	Step 1	**Step 2**
Su	.18 (.08)*	.11 (.10)
PV	.67 (.10)***	.57 (.12)***
Co	−.04 (.08)	−.17 (.13)
In		.30 (.23)
R^2	.676	.690

Note. Entries are unstandardized b*s and* SE*s*

Table 2: Regression models (Study 1)

The bottom panel of Table 2 shows the results of parallel regressions predicting importance. Comparably to the results of Table 1, adding interest has little predictive power beyond the other predictors. In this case, however, it is practical value that is doing nearly all of the predictive work: A 1 point increase on practical value is associated with a 0.67 point increase in importance. Surprise was weakly predictive in the Step 1, but not the Step 2, model. Overall, these variables predicted about 68% of the variance in perceived importance across items.

	Year 4	**Year 7**
In	.17	.19
Im	−.02	−.08
Su	−.15	−.12
SI	−.21	−.25
PV	−.03	−.06
Co	.35*	.38*

Note. Entries are first-order correlations with citations (square-root transformed) approximately 4 years and 7 years post-publication.

Table 3: Correlations with citations (Study 1)

Predictors of citation count. Table 3 presents the first-order correlations between citation count 4 and 7 years post-publication (square-root transformed) and the six measures collected in Study 1. Various regression specifications produce similar conclusions, so we focus here on the simple correlations as they avoid the multicollinearity issues mentioned above.

At both time points, neither interest nor perceived importance significantly predict citation counts, nor did judgments of surprise, scientific impact, or practical value. The only significant predictor was comprehensibility, $r(38) = .35$, $p = .028$ and $r(38) = .38$, $p = .017$ at 4 and 7 years, respectively.

Discussion

Several results pop out in these data. First, judgments of interest and importance are fairly independent: They depend on different factors and do not predict one another once one adjusts for those other factors. Comprehensibility was the most important guide to interest, but had no impact on perceived importance (see Oppenheimer, 2006 for related findings). Practical value was the most important determinant of perceived importance, and also had a large effect on interest. Surprisingness was correlated with interest but not perceived importance. Second, these judgments had little predictive power for citation rates, either in the shorter- or longer-term. Comprehensibility had a moderately high correlation with citations, but no other factor did.

Study 2

The Study 1 results could very well be specific to psychology. For instance, people have much more detailed intuitive theories of psychology, since they can introspect about their own psychology, and therefore surprisingness could be seen as an especially strong cue to scientific impact. Study 2 repeated this procedure on natural science findings from *Science* magazine.

Method

We recruited 60 participants from Mechanical Turk. Participants were excluded using the same criterion as Study 1 ($N = 14$).

The materials were the "editor's summaries" of 40 articles published in the January 6, January 13, and January 20, 2012 issues of *Science* magazine. These summaries are written by the editorial staff of the journal, rather than by us, eliminating the possibility of experimenter bias. We lightly edited the summaries to match the format of our Study 1 materials (replacing the authors' names with "Researchers"). For example, the editor's summary of one article (Fermi LAT Collaboration, 2012) read:

> Binary star systems that contain a neutron star or a black hole are expected to emit gamma rays. These gamma-ray binaries are a rare class of objects, which are also expected to emit x-rays. Indeed, several such systems were initially detected through their x-ray emission. Researchers have reported the detection of a gamma-ray binary that was previously unknown as an x-ray source. Follow-up observations reveal that the system is also a source of x-rays and that the companion star is a class O star, a type that is very hot and very luminous.

Participants read 10 of the 40 descriptions

(counterbalanced across participants), making the same six judgments as in Study 1, using a similar procedure.

Citation counts were obtained using the same procedure as Study 1, on 17 August 2016 and 30 January 2019.

Results

Most of the key results from Study 1 were replicated. Surprisingness, practical value, and comprehensibility all predicted judgments of interest, while only practical value robustly predicted judgments of importance. Citations were marginally predicted by comprehensibility, as in Study 1, but also by judgments of practical value.

Differences in means across studies. Table 4 presents the descriptive statistics for each judgment across each set of summaries. We compared the means on each measure across studies, using the false discovery rate procedure to adjust p-values for multiple comparisons (Benjamini & Hochberg, 1995). Overall, the natural science findings in Study 2 were viewed as less interesting than the psychology findings, $t(78) = 3.51$, $p = .002$, $d = 0.78$, 95% CI[0.41,1.48], but as more important than the psychology findings, $t(78) = 2.82$, $p = .009$, $d = 0.63$, 95% CI[0.18,1.06]. The natural science findings were also viewed as more surprising, $t(78) = 2.68$, $p = .011$, $d = 0.60$, 95% CI[0.15,1.04], and more scientifically impactful, $t(78) = 3.84$, $p < .001$, $d = 0.86$, 95% CI[0.37,1.15], but of similar practical value, $t(78) = 0.67$, $p = .51$, $d = 0.15$, 95% CI[–0.33,0.66]. Finally, the psychology findings were much easier to understand, $t(78) = 9.39$, $p < .001$, $d = 2.10$, 95% CI[2.52,3.87].

	Study 1 (Psychology)	Study 2 (Natural Science)
In	5.88 (0.84)	4.93 (1.48)
Im	5.74 (0.93)	6.36 (1.02)
Su	4.26 (1.19)	4.86 (0.75)
SI	5.35 (0.86)	6.11 (0.91)
PV	5.50 (1.00)	5.67 (1.20)
Co	7.80 (1.08)	4.60 (1.86)

Note. Entries are means (SDs) across items.

Table 4: Descriptive Statistics across Studies

First-order correlations. Table 5 shows the first-order correlations for Study 2, analogous to Table 1.

Like Study 1, there was a significant correlation between interest and importance, $r(38) = .39$, $p = .013$, although of more modest magnitude. The correlation between surprise and perceived scientific impact was also more modest. This weaker correlation, relative to Study 1, may have resulted from participants' lesser ability to rely on intuitive theories of the natural sciences than of psychology, given introspective access to one's own mental states.

	In	Im	Su	SI	PV
Im	.39*	—			
Su	.35*	.18	—		
SI	.40*	.87***	.27º	—	
PV	.45**	.76***	–.02	.72***	—
Co	.89***	.22	.18	.18	.35*

Table 5: First-order correlations (Study 2).

Predictors of interest and importance. Table 6 shows the regression coefficients predicting interest and importance judgments, analogously to Table 2.

The results are similar to Study 1. For interest judgments, we find that surprise, practical value, and comprehensibility are all significant predictors, with comprehensibility the strongest predictor, just like Study 1. (However, surprise was a stronger predictor than practical value in Study 1, whereas the converse was true in Study 2.) Like Study 1, importance does not have any added predictive power; in this case, its collinearity with practical value leads both to be non-significant when entered simultaneously.

	DV: Interest	
	Step 1	Step 2
Su	.41 (.13)**	.36 (.14)*
PV	.21 (.08)*	.09 (.13)
Co	.63 (.05)***	.64 (.06)***
Im		.18 (.15)
R^2	.854	.859

	DV: Importance	
	Step 1	Step 2
Su	.29 (.14)*	.21 (.16)
PV	.69 (.09)***	.64 (.10)***
Co	–.06 (.06)	–.19 (.13)
In		.21 (.18)
R^2	.631	.645

Note. Entries are unstandardized bs and SEs

Table 6: Regression models (Study 2)

For importance judgments, we find, just as in Study 1, that the key predictor is practical value, with a possible additional role for surprise. Given the high correlation

between scientific impact and practical value in Study 2, however (see Table 5), replacing surprise with scientific impact in the regression leads to a reversal of the coefficient magnitudes: Scientific impact is then a more robust predictor of importance than practical value, although both are significant in either model. (This is not true for Study 1, where scientific impact and surprise are basically interchangeable in the models.)

	Year 4	Year 7
In	.16	.18
Im	.16	.18
Su	−.23	−.25
SI	.03	.04
PV	.43**	.45**
Co	.27º	.28º

Note. Entries are first-order correlations with citations (square-root transformed) approximately 4 years and 7 years post-publication.

Table 7: Correlations with citations (Study 2)

Predictors of citation count. Table 7 presents the correlations of our six judgment variables with citation counts approximately 4 and 7 years post-publication. Comprehensibility was a marginally significant predictor at both time points, $r(38) = .27$, $p = .098$ and $r(38) = .28$, $p = .076$, which is consistent with the predictive power of comprehensibility for citations in Study 1. Unlike Study 1, however, practical value also predicted citation counts at both time points, $r(38) = .43$, $p = .006$ and $r(38) = .45$, $p = .004$. Thus, laypeople do appear to be able to extract some information that is predictive of the academic impact of scientific findings, but it is not necessarily reflected in their own judgments of importance.

Discussion

Study 2 replicated the main results of Study 1: Comprehensibility was a powerful cue to interest but not importance, although only a marginal predictor of citations in Study 2. Surprisingness was only a robust predictor of interest, but not importance, while practical value strongly predicted both. Unlike Study 1, practical value was a fairly strong predictor of citations, even though perceived importance was not.

General Discussion

Lay intuitions about scientific importance are, well, important. They impact our choices of research topics indirectly, as we try to appeal to laypeople's interests, and directly, as we all have a layperson inside of us (Goldberg & Thompson-Schill, 2009). What scientific findings do laypeople consider interesting and important? How much do these judgments track objective scientific importance?

Overall, comprehensibility is the most important predictor of interest. It is unclear whether this is because writing quality itself provokes interest, or because interesting findings are easier to explain clearly—quite possibly both. Scientists who wish to appeal to public interest ought to keep this demand for clarity in mind, rather than obscuring their work in jargon (see Oppenheimer, 2006).

Perceived practical value was the most robust predictor of importance, although perceived scientific impact was also highly predictive in Study 2 (and difficult to distinguish from practical value). Surprisingness appears to be less predictive. This is, ironically, quite a counterintuitive result! Guides to doing "interesting research" (Gray & Wegner, 2013) and our professional intuitions point to the importance of surprising the reader. But laypeople may well be growing weary of surprising findings, as they encounter increasing levels of "click bait" reporting and all-too-frequent reversals of conventional wisdom (are we, or are we not, supposed to eat eggs now?). Future research might investigate the factors underpinning and moderating this relationship between surprise (e.g., Maguire, Maguire, & Keane, 2011) and judgments of interest and importance.

Finally, these results suggest caution regarding our ability to predict the impact of scientific research based on its relationship with our intuitive theories. Surprise had no impact on citations, but neither did interest or judgments of importance or scientific impact. The only factors impacting citation were comprehensibility (in both studies) and perceived practical value (in Study 2).

It is important to understand how laypeople think about science because scientific progress tracks social priorities—scientists serve at the pleasure of society. To the extent that laypeople have systematic misconceptions about science, we must understand how those misconceptions might thwart the dissemination of science to the public, or even scientific progress itself. To the extent that laypeople have irreducible preferences over the kind of science they like to see, we must understand how those preferences might be reflected in the kind of research produced by scientific institutions.

Several other research programs contribute to this broad goal. For example, people favor reductionist explanations (e.g., referring to smaller parts or component processes) even when the reductionist information makes no logical contribution to the explanation (e.g., Hopkins, Weisberg, & Taylor, 2016; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). As a second example, people have consistent intuitions about the limits of science, particularly of psychology (Gottlieb & Lombrozo, 2018), believing that phenomena are scientifically explainable to the extent that scientists can make falsifiable and reductionist claims about those phenomena (Johnson, Kim, & Keil, 2016).

These research areas—characterizing what scientific explanations people find compelling and what scientific questions people find tractable—are valuable because they contribute to our understanding of how the lay public

interfaces with the scientific community. If people have an unjustified preference for neuroscientific explanations or an ill-founded belief that psychological phenomena are beyond the limits of science to comprehend, these may lead to society-wide distortions in our scientific priorities.

Our work complements these approaches. While these other lines of research hint at what the public's scientific priorities might be by characterizing folk scientific beliefs, the current studies take a more direct approach by asking what research people find interesting and important. If we believe that laypeople's standards (e.g., regarding practically valuable findings as more important) are reasonable, then this is all to the better. To the extent that we find lay preferences more questionable (e.g., favoring counterintuitive findings as more interesting), this should catalyze a discussion about how society prioritizes research questions, how journals select research findings, and how scientists choose research topics.

Scientists get into the business because they want to have an impact—maybe even to change the world. We may be less able than we believe to predict successfully what scientific innovations are indeed important. Doing so successfully may require us to step back and reconsider our habits of thought.

References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289–300.

Fermi LAT Collaboration (2012). Periodic Emission from the Gamma-Ray Binary 1FGL J1018.6–5856. *Science*, *335*, 189–193.

Frankenstein, J., Mohler, B.J., Bülthoff, H.H., & Meilinger, T. (2012). Is the map in our head oriented north? *Psychological Science*, 23, 120–125.

Goldberg, R.F., & Thompson-Schill, S.L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*, 480–487.

Gottlieb, S., & Lombrozo, T. (2018). Can science explain the human mind? Intuitive judgments about the limits of science. *Psychological Science*, *29*, 121–130.

Gray, K., & Wegner, D. M. (2013). Six guidelines for interesting research. *Perspectives on Psychological Science*, *8*, 549–553.

Harzing, A., & Alakangas, S. (2016). Google Scholar, Scopus, and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*, 787–804.

Hopkins, E.J., Weisberg, D.S., & Taylor, J.C.V. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, *155*, 67–76.

Johnson, S.G.B., Kim, K., & Keil, F.C. (2016). The determinants of knowability. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive thought* (pp. 49–81). New York, NY: Cambridge University Press.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Maguire, R., Maguire, P., & Keane, M.T. (2011). Making sense of surprise: An investigation of the factors influencing surprise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 176-186.

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and mystery: The cultural selection of minimally counterintuitive narratives. *Cognitive Science*, *30*, 531–553.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716.

Oppenheimer, D.M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, *20*, 139–156.

Popper, K. (1959). *The logic of scientific discovery*. London, UK: Routledge. (Original work published 1934.)

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215.

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*, 470–477.