Title
Perceptions of Explanation Completeness Help Decrease Knowledge Overestimation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Wilson, Julianne
Marsh, Jessecae K.

Publication Date
2023

Peer reviewed

Perceptions of Explanation Completeness Help Decrease Knowledge Overestimation

Julianne Wilson (juw521@lehigh.edu)
Department of Psychology, 27 Memorial Dr W.
Bethlehem, PA 18015

Jessecae K. Marsh (jessecae.marsh@lehigh.edu)
Department of Psychology, 27 Memorial Dr W.
Bethlehem, PA 18015

Abstract

The tendency to overestimate one's knowledge has been shown in many domains including the innerworkings of everyday objects. This Illusion of Explanatory Depth (IOED) can be broken through the act of generating a causal explanation, although the reason as to why has yet to be explored. In this study, we investigate what characteristics of a generated explanation result in people recognizing their perceived lack of knowledge. Participants completed a typical IOED paradigm for devices, followed by rating their perceived completeness and accuracy for the explanations they generated. We also coded the explanations to determine their causal complexity. We found that lower ratings of overall perceived completeness and a sense of incomplete big explanatory components were predictive of a larger decrease in perceived understanding for that device post-explanation. Fewer causal links within an explanation also predicted a larger decrease in understanding ratings, suggesting that producing an explanation with a lower causal complexity led to a decrease in perceived understanding of that device. We discuss the implications of these results in relation to explanation characteristics that may cause a person's illusion of understanding to break and proposed origins of the IOED phenomenon.

Keywords: Illusion of explanatory depth; causal relationships; causal knowledge; explanation

Introduction

Almost everyone has used a can opener at some point in their lives. Because a can opener is such a familiar and commonly used object, people may also believe they have a strong understanding of how it works. However, if they were asked to write a detailed, step-by-step causal explanation as to how a can opener operates, they may realize their understanding is not as complete as they once thought. In line with this hypothetical, multiple studies have shown that people's causal understanding can be exceedingly shallow and filled with holes (Matute et al., 2015; Wilson & Keil, 1998). In addition, people are often unaware of their lack of understanding, leading to a phenomenon known as the Illusion of Explanatory Depth (IOED; Rozenblit & Keil, 2002). This illusion of understanding has been shown to break through the generation of a causal explanation about the subject of interest, such as how does a can opener open a can. While the exposure of the IOED through generating a causal explanation has been repeatedly shown (see e.g., Fernbach et al., 2013; Lawson, 2006; Vitriol & Marsh, 2021; Zeveney & Marsh, 2016), little is known about why this is the case – i.e., what in the explanations people generate results in their readjustment of perceived understanding. In this research, we explore this question, namely, what in the content of a generated explanation leads to people breaking their illusion of understanding.

Rozenblit and Keil (2002) developed a paradigm using explanation generation to show both the existence and the ability to break the IOED in people's understanding of household devices and natural phenomena. The basic IOED paradigm works as follows: All participants are first presented with detailed instructions on how to assess their own understanding of different objects using a 1 to 7 scale, in addition to an example of what a low (1), middle-of-the-road (4), and high (7) rating explanation would look like. They are then asked to rate their understanding of a list of topics using the same 1 to 7 scale they just learned. Next, participants are asked to physically write out a step-by-step causal explanation for one of the initially rated objects, immediately followed by being asked to re-rate their understanding of that same topic. The cycle of explanation generation and re-rating continues until all selected topics have been completed. Traditional analyses average all pre-explanation ratings and all post-explanation ratings across items and compare across the two time points. If post-explanation ratings are found to be significantly lower than pre-explanation ratings, this suggests that an IOED was both present (due to the higher pre-explanation ratings) and successfully broken (due to the lower post-explanation ratings) for the explained items.

Although additional work using this paradigm has since shown the IOED to exist and be breakable through the generation of explanations in numerous domains including complex devices (Johnson et al., 2016; Lawson, 2006), politics (Fernbach et al., 2013), mental health (Zeveney & Marsh, 2016), and historical knowledge (Gaviria & Corredor, 2021), the characteristics of the explanations themselves that lead to this knowledge reassessment are still unknown. In our study, we focused on three aspects of an explanation that may influence perceived understanding judgments: perceived completeness, perceived accuracy, and causal complexity.

A person may adjust their perceived understanding of a subject based on the level of completeness of the self-generated explanation they are able to produce. For example, while generating an explanation, one may recognize that there are large gaps of information missing that they are unable to recall, displaying the incompleteness of their knowledge (Keil, 2006). Completeness is also considered a

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

component of explanatory coherence (Trout, 2002), a virtue of explanations that is used in determining the quality of an explanation. Therefore, holes in a causal explanation may decrease its value (Zemla et al., 2017). People rate information that contains extended elaborations as having a higher importance in the overall value of the explanation, compared to information that does not have the same elaborations or information that has elaborations that are too technical to comprehend. This was specifically found for mechanistic information and not general filler information (Rottman & Keil, 2011). These findings provide support for the idea that completeness of some form is important to perceiving an explanation as high quality. While Rottman and Keil's (2011) findings suggest the importance of completeness, they do not specify what components specifically need to be present to be elaboratively complete. Zemla et al. (2017) proposed that the level of detail in an explanation could be important to perceptions of the explanation's adequacy. It is therefore an open question as to what must be elaborated on, i.e., are major components of importance enough or must small details be given as well to provide a sense of an explanation being complete.

Another aspect of a self-generated explanation that could influence perceived understanding judgments is the perceived accuracy of generated components. To alter one's conclusion in an argument, a person must believe there is an error in their reasoning or that an alternate reasoning is of better quality (Allen & Burrell, 1992). A person perceiving their explanation is inaccurate or has flawed reasoning could lead them to reconsider their understanding of the topic at hand. This is why regardless of whether people are truly accurate (see Griffin et al., 2009), we are interested in how their perceptions of that accuracy influence their perceived understanding. People may also be able to generate many steps in a causal explanation that could make it look complete, but not be sure of the actual accuracy of each component. This may, in turn, expose their lack of understanding of the subject and lead them to decrease their understanding ratings separately from how complete they deem their explanation to be.

People could also adjust perceptions of their own understanding in accordance with the causal complexity of the explanation they generate. An explanation may be considered complex for a number of reasons, from its structure to its contents. We operationalized causal complexity by identifying the number of individual causal links present, similar to how Zemla et al. (2017) measured causal complexity. Our reasoning is that, if a person is able to generate more causal links in total, regardless of what those links were, they may have a stronger sense of their own ability to understand the phenomenon. This allows us to separate causal complexity from accuracy with the focus being on simply the generation of links, regardless of whether or not they are accurate.

In the following experiment, we explore how people's perception of the completeness and accuracy of their generated explanations, as well as the number of causal links in the explanations, relates to their perceived understanding of what they explained. Each of these characteristics could lead someone to reassess their knowledge on a particular topic and potentially lower their understanding rating depending on their conclusion.

Method

Participants

Participants were 51 undergraduate students at Lehigh University compensated with credit toward the research participation requirement in their introduction to psychology course. Participants were fluent English speakers and had normal or corrected-to-normal vision.[1] Participants who could not meaningfully answer our screening questions were excluded from analysis (*n* = 1).

Materials

We selected ten devices from those used by Rozenblit and Keil (2002) for their original IOED experiment. Five devices served as stimuli to be explained (how a can opener works, how piano keys make sound, how a car ignition system starts an engine, how a flush toilet operates, how a zipper works) and 5 would not be explained and served as filler items, as in Rozenblit and Keil (how a water faucet controls water flow, how a ballpoint pen writes, how a sewing machine works, how a helicopter flies, how a spray bottle sprays liquids).

We based our instructions on the instructions used in Rozenblit and Keil (2002) to inform participants of how to rate their understanding of an item on a 1 to 7 scale where higher ratings reflected greater understanding of the phenomenon. The instructions included an example of an explanation warranting the lowest possible score (1), a midrange score (4), and the highest possible score (7).

Measures

IOED Understanding Measure Participants completed the same time 1 (T1) and time 2 (T2) measures described in Rozenblit and Keil (2002; Experiments 1 through 4, phases 1 through 3). The T1 prompt stated, "For each of the following, please rate your understanding using the 1 to 7 scale that you just learned about." The T2 prompt was presented as "Now, please rate how well you feel you understand X," with "X" being replaced with the devices listed above. Responses for both time ratings were made on a 1 (Very vague understanding) to 7 (Very thorough understanding) scale.

Explanation Prompt Adapted from Rozenblit and Keil (2002), participants were prompted to make their explanation

[1] An additional 51 participants were recruited and completed a version of the experiment that did not have a traditional explanation prompt (i.e., a more descriptive prompt). Those data are not reported here. Participants were randomly assigned to complete the traditional IOED version that is reported, or the alternative version that is not reported.

as follows: “Now, we’d like to probe your knowledge in a little more detail on some of the items. As best you can, please describe all the details you know about X, going from the first step to the last, and providing the causal links between the steps. That is, your explanation should state precisely how each step causes the next step in one continuous chain from start to finish. In other words, try to tell as complete a story as you can, with no gaps. Please take your time, as we expect your best explanation,” with “X” being one of the five devices listed above.

Perceived Completeness We asked three questions that assessed participants’ perceptions of the completeness of their written explanations. First, participants were asked to estimate how much of the possible information they generated (% Complete) as follows: “Think about everything a person could have produced in generating an explanation of X. What percent of the possible information do you think you produced?” (sliding scale from 0% to 100%).

Next, participants rated the completeness of certain parts of the explanation they generated. Specifically, we asked their agreement on a 1 (strongly disagree) to 7 (strongly agree) scale of whether they included important parts of the explanation (Big Parts Inclusion: “I included all of the big, important parts that would need to be in an explanation of X.”), as well as small details (Small Parts Inclusion: “I included all of the small, less important details that could be in an explanation of X.”). Again, in all cases “X” was substituted for one of the 5 explained devices.

Perceived Accuracy We also measured how accurate participants perceived their explanations to be regardless of how complete they were. We developed three questions that measured perceived accuracy and paralleled the completeness questions. First, we asked participants to estimate the accuracy of the information they specifically produced (% Accuracy): “Think about everything you produced in your explanation of X. What percent of the information you *actually* generated do you think was accurate?” (sliding scale from 0% to 100).

Participants were then asked to rate their agreement with the accuracy of the important parts of the explanation they included (Big Parts Accuracy: “Please select how much you agree or disagree with the following statement based on the accuracy of your explanation of X: All of the big, important parts I included were correct.”) as well as the small details they included (Small Parts Accuracy: “Please select how much you agree or disagree with the following statement based on the accuracy of your explanation of X: All of the small, less important details I included were correct.”) on a 1 (strongly disagree) to 7 (strongly agree) scale. Again, in all cases “X” was substituted for one of the 5 explained devices.

Causal Complexity Coding Each participant explanation was coded by two independent coders for the number of parts present and the number of causal links present. A part described some part of the device. A causal link was defined as the presence or inference of a part acting on another part. For example, in the phrase “The user clamps down firmly on the handles,” there is one causal link with “user” distinguished as the part acting, “clamps down” as the action, and “handles” as the part being acted on. After coding independently, the coding pairs met to settle all differences by discussion. Because of the high correlation between number of parts and number of causal links, we use only causal links as a measure of causal complexity in the analyses. Coders also coded if participants used analogies in their explanation and whether they expressed open uncertainty. These codes were infrequently applied and so they were not analyzed.

Demographics and Screening Questions Participants were asked general demographic questions, along with two screening questions to test their understanding of the experiment. The screening questions included the following: “What was the current study about?” and “Please describe what you did during the study.”

Procedure

Participants performed the experiment in-person on a lab-provided computer. After providing consent, participants read the instructions for rating their understanding. Participants were then asked to use the newly-learned scale to rate their understanding of the ten devices (T1 ratings), which included 5 test items for which explanations would be generated and 5 filler items. Next, participants generated explanations for the 5 test items they just rated. After explaining one device, they immediately re-rated their understanding of that test item (T2 rating). They repeated explaining and then re-rating for the remaining test items. The order of the items displayed during the T1 rating and the order in which test items were asked to be explained and re-rated were randomized for each participant.

Participants then answered the perceived completeness and perceived accuracy questions (six questions for each device) without being able to look back at their generated explanations. All completeness and accuracy questions for a given device were asked as a set in the same order described above. The order of devices was randomized for each participant. Lastly, participants completed the demographic questions and screening questions.

Results

Changes in Perceived Understanding Ratings

Our main goal in these analyses is to test how people’s subjective impressions of the explanations they generated and the causal complexity of those explanations are related to decreases in perceived understanding ratings seen in the IOED paradigm. To that end, we first tested whether the IOED effect replicated in these data. To do this, we performed the traditional analysis conducted in this literature. Namely, we averaged understanding ratings across devices at both T1 and T2 for explained devices and submitted those means to a one-way ANOVA with Time (T1 vs. T2) as the main factor. Average T2 ratings (M = 3.54, SE = 0.062) were significantly lower than average T1 ratings (M = 4.23, SE = 0.077), $F(1, 254) = 234.2$, $p < .001$, $\eta_p^2 = .48$ (Figure 1). These results are in line with previous research on the IOED showing that understanding ratings significantly decrease after participants write out an explanation of a device.

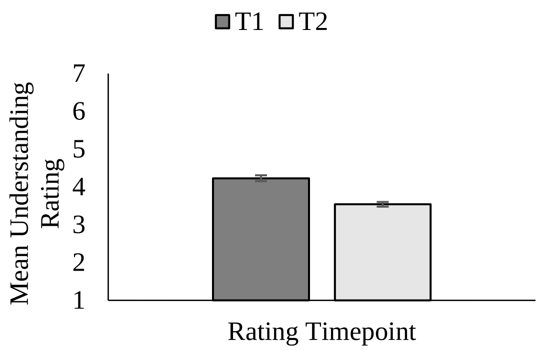

Figure 1: The average understanding ratings at T1 and T2. Error bars indicate standard error.

Given that our goal is to determine what in an explanation drives a person to reassess their understanding, we next explored whether there was variation across participants in whether they decreased their ratings after generating an explanation. As can be seen in Figure 2, there was variability across devices in relation to where participants initially rate their understanding.[2] Because this variability leaves more or less room to significantly decrease T2 ratings, we looked within each device at what percentage of participants decreased, had no change, or increased their understanding ratings from T1 to T2. Table 1 shows that across devices, a larger percentage of participants decreased their ratings post-explanation than either remained the same or increased their ratings. We next turn to analyses that help us understand what in a generated explanation leads a given person to decrease their perceived understanding ratings for a specific device.

Completeness, Accuracy, and Complexity

The main purpose of this experiment was to determine if perceived completeness, accuracy, and/or the causal complexity of a generated explanation was predictive of a change in perceived understanding ratings. To do this, we used linear mixed modelling (LMM). We included in the model all 6 perceived completeness and perceived accuracy measures, as well as the number of causal links, for a total of 7 predictor variables. Additionally, the participant-level average of the T1 and T2 ratings for explained items was added to the model as a covariate to account for differences in overall base understanding ratings by participant. All predictor variables and covariates were centered on their grand mean before being added to the model. Device was entered as a repeated measure and both subjects and the average of the T1 and T2 ratings covariate were entered as random effects. Change in understanding rating, or change score (CS), was calculated by subtracting T1 ratings from T2 ratings for each explained device and included as the dependent measure. This method of calculating CS means that larger decreases are represented as larger negative numbers. The structure of the model with the best fit, determined by the lowest Akaike Information Criterion (AIC), was found to be compound symmetry heterogeneous.

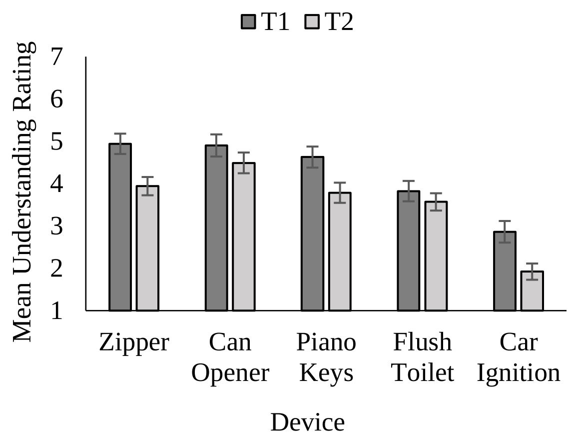

Figure 2: Average understanding ratings at T1 and T2 by device. Error bars indicate standard error.

Table 1: Percentages of participants with change in understanding ratings post-explanation per device.

Device	Decrease	No Change	Increase
Zipper	64.7%	19.6%	15.7%
Can opener	41.2%	39.2%	19.6%
Piano keys	58.8%	29.4%	11.8%
Flush toilet	39.2%	29.4%	31.4%
Car ignition	56.9%	39.2%	3.92%

[2] A 2 (time: T1 vs T2) x 5 (device: zipper, can opener, piano keys, flush toilet, car ignition) repeated-measures ANOVA showed significant main effects and a significant interaction. All devices had a significant decrease from T1 to T2 ($ps < .01$) except flush toilet ($p = .180$).

Table 2: Descriptive Statistics and Pearson Correlations.

Predictor	Mean	SE	1	2	3	4	5	6
1. % Complete	42.9	1.77	1.00					
2. Big Parts Inclusion	4.52	0.12	.78	1.00				
3. Small Parts Inclusion	2.93	0.11	.63	.48	1.00			
4. % Accuracy	59.2	1.97	.65	.66	.35	1.00		
5. Big Parts Accuracy	4.90	0.10	.62	.80	.38	.77	1.00	
6. Small Parts Accuracy	3.88	0.10	.57	.52	.64	.59	.61	1.00
7. Causal Links	4.36	0.16	.33	.38	.20	.29	.30	.24

*All correlations *ps* < .001

Predictor variables (descriptive statistics and correlations found in Table 2) were all added to the same model to decrease the risk of Type I errors due to significant correlations and/or multicollinearity. High positive correlations were seen between Big Parts Inclusion and % Completeness, Big Parts Inclusion and Big Parts Accuracy, and with Big Parts Accuracy and % Accuracy.

Table 3 shows the results of the LMM analysis. Within the model, 3 of the 7 predictor variables were found to be significant: Percent Complete, Big Parts inclusion, and Causal Links. The model was then re-run only including the three variables found to be significant as predictors, with the rest of the model kept the same (including the T1 and T2 average ratings as a covariate). All three predictor variables were again found to be significant.

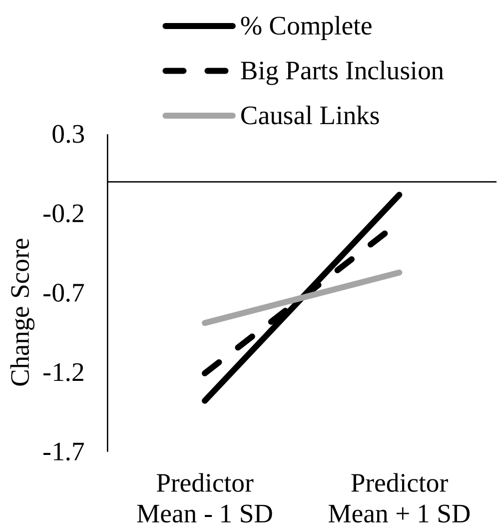

Figure 3: The effect of significant predictors on the difference in understanding rating from T1 to T2.

Figure 3 provides a visualization of how the three significant predictors related to change in understanding scores. The LMM results of the three-predictor model analysis were used to calculate how much change in understanding (e.g., Change score) was reflected for a participant whose values for each predictor were at one standard deviation below and above the mean for each of the three predictor variables. Lower points on the y-axis of the graph show a greater decrease in understanding post-explanation. As can be seen in Figure 3, each of the three predictors positively predicts the magnitude of difference in understanding ratings. Specifically, participants with lower percent complete and big part inclusion ratings, and less causal links in their generated explanations had a larger decrease in understanding ratings post-explanation.

Discussion

Numerous studies have shown people to have inaccurate perceptions of their knowledge (e.g., Matute et al., 2015; Wilson & Keil, 1998; Rozenblit & Keil, 2002; Zeveney & Marsh, 2016) and superficial causal knowledge (Sloman et al., 2021; Rabb et al., 2019) through the use of the IOED paradigm. We add to this literature by exploring what in the nature of an explanation a person generates makes them likely to decrease ratings of their perceived understanding. We found that perceptions of overall completeness and how many big, important parts of the explanation were included were related to a reduction of perceived understanding. That is, believing that there were missing parts and that those parts were not small details, but large important elements of the explanation were predictive of decreased understanding ratings. Interestingly, perceived accuracy of generated explanations did not relate to changes in understanding ratings. Importantly, our coding analysis suggests that our measure of causal complexity, namely the number of causal links generated in an explanation, was also predictive of a reduction in perceived understanding. The less causal links a person generated, the more likely they were to lower their understanding ratings.

This is the first work to measure what aspects of self-generated explanations relate to an internal recalibration of perceived causal understanding. Our findings suggest that people believe that to understand something involves the ability to produce a complete report of all important causal elements, and are less concerned about small mechanistic details. These results add to the findings of Rottman and Keil (2011) and Zemla et al. (2017) by suggesting that participants

Table 3: LMM Results.

Predictor	Estimate	SE	95% CI Lower	95% CI Upper	df	*t*	*p*
Seven Predictor Model							
% Complete	0.022	0.005	0.012	0.033	221.9	4.11	**< .001**
Big Parts Inclusion	0.177	0.080	0.020	0.334	205.3	2.22	**.028**
Small Parts Inclusion	-0.020	0.064	-0.145	0.105	224.5	-0.31	.751
% Accuracy	0.038	0.066	-0.092	0.168	194.8	0.58	.564
Big Parts Accuracy	0.001	0.004	-0.006	0.008	167.9	0.26	.796
Small Parts Accuracy	0.094	0.085	-0.075	0.262	195.3	1.10	.274
Causal Links	0.063	0.030	0.003	0.122	211.6	2.07	**.039**
Three Predictor Model							
% Complete	0.023	0.005	0.014	0.032	212.7	4.87	**< .001**
Big Parts Inclusion	0.258	0.062	0.136	0.380	158.6	4.19	**< .001**
Causal Links	0.063	0.030	0.004	0.122	209.1	2.11	**.036**

deem major components to be enough within an explanation to add explanatory value and be considered complete, even without small details.

The lack of the influence of accuracy on post-explanation ratings could suggest that people are more concerned with having a complete overview of a causal process as opposed to the accuracy of each component. One explanation for this is that people may have a bias to believe anything they generated was accurate and therefore do not use perceived accuracy as a gauge of understanding. This is partially supported by the high average overall Percent Accuracy ratings (see Table 2). Future research could explore if the overall overconfidence in causal understanding replicates for the individual pieces of information people generate in an explanation and how accurate they seem.

We found that people who produced fewer causal links overall were more likely to reduce their perceived understanding ratings. This count of causal links gives an overview of causal complexity much like Percent Complete and Percent Accuracy offer a view of overall perceived completeness and accuracy, respectively. We did not attempt to code "important" parts within explanations to match our measures of perceived big and small parts. We cannot begin to assume we can independently identify what participants would personally determine "important" in their generated explanations. It is an open question as to how a person who does not understand how a device like a can opener works comes to believe any given causal link they generate is an "important" component. Further research can explore how novices in a domain form impressions of what is important in a causal explanation.

A limitation of our findings is that we do not know for sure the direction of influence in our variables. We modeled self-perceived understanding drops as being predicted by perceived completeness of explanations. Alternatively, if people have realized they did not understand how a device worked, they could subsequently rate their completeness and accuracy as lower. If this was the direction of causality, then it would still be an open question as to what was drives lowering of understanding ratings. While we think this direction of influence is unlikely, it is an avenue for further research.

Our findings help differentiate explanations for why the IOED phenomenon exists in the first place. One explanation for the IOED is a "regression toward the mean" account; that is, rating things multiple times results in more conservative ratings. If the IOED was the result of regression to the mean, we would not expect something like perceived completeness or actual number of causal links to predict drops in understanding. An alternative hypothesis for the IOED comes from the community of knowledge approach (Fernbach & Light, 2020; Sloman & Rabb, 2016). The community of knowledge hypothesis points to a type of cognitive offloading where the causal narratives within people's minds can include placeholders as opposed to actual knowledge (Rabb et al., 2019; Sloman et al., 2021). To make the community of knowledge account fit our findings, we would suggest that people may specifically confuse experts' knowledge of big explanatory elements with their own. Given that small details did not influence changes in understanding ratings, then these elements may not be confused with the community of knowledge. For example, knowledge of small details (e.g., specific jargon or rare details) may be particularly indicative of expert knowledge and therefore harder to confuse with a layperson's own knowledge. This is very speculative, but suggests an area of future research.

The goal of this study was to explore how the perceived completeness and accuracy, as well as the causal complexity of generated explanations influence people to recalibrate their understanding. We found that people are sensitive to whether they missed large explanatory components and that this along with actually generating less links predicts downgrading their knowledge. These results suggest completeness is an important factor in whether someone is able to more precisely distinguish their own knowledge from what they perceived it to be.

References

Allen, M., & Burrell, N. (1992). Evaluating The believability of sequential arguments. *Argumentation and Advocacy*, *28*(3), 135–144. https://doi.org/10.1080/00028533.1992.11951539

Fernbach, P. M., & Light, N. (2020). Knowledge is shared. *Psychological Inquiry*, *31*(1), 26–28.

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*(6), 939–946. https://doi.org/10.1177/0956797612464058

Gaviria, C., & Corredor, J. (2021). Illusion of explanatory depth and social desirability of historical knowledge. *Metacognition and Learning*, *16*(3), 801–832. https://doi.org/10.1007/s11409-021-09267-7

Griffin, T. D., Jee, B. D., Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition, 37*(7), 1001-1013.

Johnson, D., Murphy, M., & Messer, R. (2016). Reflecting on explanatory ability: A mechanism for detecting gaps in causal knowledge. *Journal of Experimental Psychology: General*, 573–588. https://doi.org/10.1037/xge0000161.supp

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*(1), 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Lawson, R. (2006). The science of cycology: Failures to understand how everyday objects work. *Memory & Cognition*, *34*(8), 1667–1675. https://doi.org/10.3758/BF03195929

Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00888

Rabb, N., Fernbach, P. M., & Sloman, S. A. (2019). Individual representation in a community of knowledge. *Trends in Cognitive Sciences*, *23*(10), 891–902. https://doi.org/10.1016/j.tics.2019.07.011

Rottman, B. M., & Keil, F. C. (2011). What matters in scientific explanations: Effects of elaboration and content. *Cognition*, *121*(3), 324–337. https://doi.org/10.1016/j.cognition.2011.08.009

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562.

Sloman, S. A., Patterson, R., & Barbey, A. K. (2021). Cognitive neuroscience meets the community of knowledge. *Frontiers in Systems Neuroscience*, *15*, 1–13. https://doi.org/10.3389/fnsys.2021.675127

Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science*, *27*(11), 1451–1460.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *The University of Chicago Press*, *69*(2), 212–233.

Vitriol, J. A., & Marsh, J. K. (2021). A pandemic of misbelief: How beliefs promote or undermine COVID-19 mitigation. *Frontiers in Political Science*, *3*, 1–14. https://doi.org/10.3389/fpos.2021.648082

Wilson, R. A., & Keil, F. (1998). The shadows and shallows of explanation. *Minds and Machines*, *8*(1), 137–159.

Zemla, J. C., Sloman, S., Bechlivanidis, C., & Lagnado, D. A. (2017). Evaluating everyday explanations. *Psychonomic Bulletin & Review*, *24*(5), 1488–1500. https://doi.org/10.3758/s13423-017-1258-z

Zeveney, A. S., & Marsh, J. K. (2016). The illusion of explanatory depth in a misunderstood field: The IOED in mental disorders. In A. Pagafragou, D. Grodner, D. Mirman & J. C. Trueswell (Eds.), Proceedings of the 38th annual conference of the cognitive science society. Austin, TX: Cognitive Science Society.