Title
The Transferability of Explanation-Induced Knowledge Reassessment

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Wilson, Julianne
Marsh, Jessecae K.

Publication Date
2024

The Transferability of Explanation-Induced Knowledge Reassessment

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

When someone realizes they do not actually know how a can opener works, do they think it is just a one-time bout of overconfidence? Or, do they assume they lack understanding of all the devices in their home? Causal knowledge is a fundamental part of both daily functioning and long-term learning. Previous studies have shown that writing out a causal explanation has the ability to induce knowledge reassessment and decrease inflated perceptions of knowledge specific to the concept being explained. However, the generalization of this knowledge reassessment has only recently been explored. In this preregistered experiment, we used the Illusion of Explanatory Depth (IOED) paradigm to see whether a decrease in perceived understanding of an explained item affects the perceived understanding of an item that was not asked to be explained. We also assessed the effect of explanation quality on this transfer of knowledge. Results showed that knowledge reassessment for explained items led to an even greater reassessment for unexplained items, suggesting possible overgeneralization. While explanation quality influenced knowledge reassessment for explained items, it did not for unexplained items. We discuss the possible reasons for these results as well as future studies to help understand the boundaries of knowledge reassessment.

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

Introduction

Humans are constantly learning how different items and events relate to one another through cause and effect. This causal knowledge serves as a guide for making inferences (Matute et al., 2015) and categorizing the world around us (Keil, 2003). Because causal understanding is so central in making judgments and real-world decisions about important aspects of daily life (Bes et al., 2012), it is tempting to believe that we have an adequate grasp of the inner workings of the world or, at least, that we are aware of what we know and what we do not know. However, numerous studies reveal the opposite – that people's causal understanding is often surface-level and gap-ridden (Matute et al., 2015; Wilson & Keil, 1998). Even worse, we tend to be unaware of our own ignorance, resulting in a phenomenon commonly referred to as the Illusion of Explanatory Depth (IOED; Rozenblit & Keil, 2002).

A paradigm developed by Rozenblit and Keil (2002) has been used by numerous studies to show that people have inaccurate perceptions of their own causal knowledge. A general version of this paradigm proceeds as follows: (1) Participants are given instructions for using a 7-point scale to rate their understanding of different concepts with example explanations for a low, middle, and high rating; (2) Participants are asked to use this scale to rate their understanding of a number of specific concepts (Time 1 [T1] ratings); (3) Participants are instructed to write a detailed causal explanation for a portion of the initially rated concepts and, after writing each explanation, are asked to re-rate their understanding of that particular concept (Time 2 [T2] ratings); (4) Explanation generation and secondary rating repeats for all concepts selected by the experimenter.

Data from this paradigm has typically been averaged across concepts at each time point, with a significant decrease from T1 to T2 signifying both the initial presence and successful reassessment of the IOED. This paradigm has given ample evidence that attempting to provide a causal explanation for certain concepts causes a reassessment of perceived knowledge for that particular concept. This has been shown in a variety of domains including devices and natural phenomena (Johnson et al., 2016; Lawson, 2006; Rozenblit & Keil, 2002), politics (Fernbach et al., 2013), mental health (Zeveney & Marsh, 2016), and historical knowledge (Gaviria & Corredor, 2021).

While the exact mechanism of the IOED remains a mystery, Rozenblit and Keil (2002) found it to be most prevalent for causal knowledge as opposed to other types of knowledge such as facts, narratives, and procedures. This suggests that its pervasiveness is caused by more than mere overconfidence (Mills & Keil, 2004). In addition, Johnson et al. (2016) showed that deeply reflecting on a causal explanation has the ability to reduce perceived knowledge ratings, but to a smaller degree than physically writing out an explanation (Experiments 1 and 5). These results stress the importance of the written explanation in the degree of knowledge reassessment after the IOED paradigm.

A recently explored question from this research is how causing someone to reassess their knowledge of one concept affects other elements of people's reasoning. In this preregistered research, we investigate whether the broken illusion of causal knowledge that occurs during explanation generation generalizes to unexplained items. To understand why this generalization may occur requires connecting the IOED and the paradigm used to test it to the broader area of metacognition.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

Metacognition, the IOED, and Knowledge Transfer

Assessing one's own knowledge, or metacognitive monitoring (Rhodes, 2019) plays a crucial role in the IOED paradigm at two timepoints - before and after the participant is asked to explain the concept. Ratings made before giving an explanation (T1 ratings) can be considered a prospective metacognitive judgment because they occur before participants are asked to give any evidence for their rating. Conversely, ratings made after explanation generation (T2 ratings) are a retrospective metacognitive judgment because they occur after participants are asked to prove their knowledge with an explanation.

Both prospective and retrospective judgments are grounded in declarative knowledge and subjective experience (Siedlecka et al., 2016). How the two judgments differ from one another depends on the introduction of any new information acquired in the time between a prospective and retrospective judgment (Koriat & Levy-Sadot, 2000). By this account, T2 ratings in the IOED can only be influenced by the portion of the paradigm that occurs between judgements - generating a causal explanation. If T2 ratings are found to be significantly lower than T1 ratings, it suggests that the relative difficulty that participants experience in generating an explanation acts as introspective feedback (Schwarz, 2004), causing a reassessment of perceived knowledge about that particular concept and resulting in a lower retrospective judgment.

Thinking of the T1 and T2 ratings of the IOED paradigm as prospective and retrospective metacognitive judgments opens up a wide range of literature with which to form hypotheses about how the IOED task may affect knowledge reassessment more broadly. For example, Carpenter et al. (2019) showed that participants who received feedback on their metacognitive accuracy – i.e., the degree of agreement between their ratings of understanding and objective accuracy – had a greater ability to improve their future metacognitive accuracy than participants that only received feedback about their objective accuracy. In addition, this increase in metacognitive accuracy was seen in tasks both similar and distinct from the task during which participants originally received the feedback. These results suggest that the addition of metacognitive feedback, much like the introspective feedback that occurs during the IOED paradigm, may allow for more generally accurate metacognitive judgments in the future.

Previous Work on Knowledge Transfer in the IOED

In a preliminary investigation, Roeder (2016) asked participants to explain one set of items and then rate their understanding of a set of unexplained items. Roeder found that carryover to unexplained items could occur, but that the decrease in understanding ratings was smaller than what occurs for explained items (Experiment 1). However, the evidence from their additional experiments did not allow for strong conclusions as they used items like procedures that did not (and have traditionally not) shown an IOED.

After completing our following experiments, we were pointed to a recently published article by Meyers et al. (2023) who explored the transferability of causal knowledge using the IOED. Meyers et al. found a decrease in post-explanation understanding ratings when compared to pre-explanation ratings for both explained and unexplained items. However, as discussed by Meyers et al., the statistical methods they used failed to consider the variability in ratings among items.[1] In addition, Meyers et al. (2023) were unable to directly compare differences in understanding ratings for explained vs. unexplained items, and even caution readers against doing so in their results (p. 7).

Overall, the work presented in Meyers et al. (2023) is a sound preliminary investigation into the generalization of knowledge transfer within the IOED paradigm that our work serves to expand upon. Previous work in our laboratory assessing explanation quality has shown that the amount of decrease in understanding ratings from pre-explanation to post-explanation had a significant negative relationship with three aspects of the quality of participants' explanations: participants' perceptions of their explanations' overall completeness, whether they felt as though their explanations were missing important details, and the number of causal links in the explanations (determined by blind coders). Namely, a larger decrease in understanding ratings from T1 to T2 was predicted by lower ratings of perceived completeness and inclusion of important details, as well as a smaller number of causal links present in explanations (Wilson & Marsh, 2023). These results show that the completeness of a participant's explanation may be a driving force in whether they reassess their knowledge, but it leaves an open question as to whether explanation quality influences knowledge transfer in a similar way. For example, are people who believe they gave poor explanations more likely to determine they know less about other, unexplained items?

Experiment Overview and Hypotheses

Our experiment investigates the generalization of knowledge reassessment from explained items to unexplained items using the IOED paradigm. Preregistration and supplementary materials for this experiment can be found at https://osf.io/8h2k5. Unlike Roeder (2016), we attempted to observe this transfer within causal knowledge only and within the same stimuli domain (i.e., household devices). In addition, we used a within-subject design, having subjects re-rate both devices they were asked to explain during the paradigm as well as devices they were not asked to explain. This allows for a more direct comparison when assessing

[1] Meyers et al. (2023) performed alternative analyses in their supplementary materials, but do not make conclusions based on these results.

differences in understanding ratings for explained and unexplained items.

Our experiment expands on the findings of Meyers et al. (2023) in three ways: 1. Methodology - participants explained multiple devices as opposed to one device, and the methodology and instructions used are akin to the original IOED paradigm; 2. Analyses - we investigated not only the possibility of knowledge transfer but also the strength of knowledge transfer by directly comparing ratings for explained and unexplained items; and 3. Influence of Explanations - we looked at whether the characteristics of the explanations as well as participants' perceptions of their explanations influenced the degree of knowledge transfer or lack-thereof.

Based on the literature discussed, we hypothesized that knowledge reassessment would successfully transfer to unexplained items. We also hypothesized that the perceived completeness, inclusion of big details, and number of causal links in participant-generated explanations would be predictive of the degree of knowledge transfer from explained to unexplained items, similar to what is seen for explained items in Wilson and Marsh (2023).

Method

Participants

Participants were 74 undergraduate students (age M = 18.97, range 18 - 21) who predominately identified as men (86%; women = 5%; nonbinary = 4%; preferred to self-describe = 1%; preferred not to respond = 3%), as White (63%; African American = 6%; Asian = 15%; preferred to self-describe = 9%; preferred not to respond = 8%), and not Hispanic (82%; Hispanic = 14%; preferred not to respond = 4%). They were enrolled in the introduction to psychology course at a northeastern private university and were compensated with research credit toward their course. Inclusion criteria were that participants were fluent in English and have normal or corrected-to-normal vision. Participants who did not provide ratings for all items were excluded from analysis (n = 3). Additionally, two participants were removed from analysis due to their responses to screening questions. The first participant explicitly expressed their lack of fluency in English while the other stated they had previous knowledge about the nature of the experiment. Power was calculated the same way as in Wilson and Marsh (2023).

Materials

We chose materials from Rozenblit and Keil's (2002) original IOED experiments. Because many of the devices used by Rozenblit and Keil can be considered outdated (i.e., a VCR), we chose eight devices that the undergraduate population were likely to be familiar with from this list (can opener, piano keys, flush toilet, zipper, spray-bottle, ballpoint pen, water faucet, and "cylinder lock" was changed to "lock" for clarity.) Four other common household devices (toaster, freezer, printer, electric blanket) were chosen to complete the list. The explanation prompts came in the generic form of "how a {insert device name} works" (e.g., how a can opener works) with the following exceptions: "how piano keys make sounds", "how a flush toilet operates", "how a spray-bottle sprays liquids", "how a ballpoint pen writes", "how a key opens a lock", and "how a water faucet controls water flow".

Initial instructions were based on the instructions used in Rozenblit and Keil (2002), teaching participants how to rate their understanding using a 1 (lower understanding) to 7 (higher understanding) scale. They were also given an example of an item explanation that fit with the lowest possible score (1), the midrange score (4), and the highest possible score (7).

Measures

IOED Understanding Measure We asked participants to rate at T1 the following understanding question: "For each of the following, please rate your understanding using the 1 to 7 scale that you just learned about." The T2 prompt for explained items varied this question to: "Now, please rate how well you feel you understand X," with "X" being replaced with the device phrases listed above (i.e., "how a can opener works"). These are the same questions as used in Rozenblit and Keil (2002). All ratings were made on a 1 (Very vague understanding) to 7 (Very thorough understanding) scale. Before making secondary ratings for the unexplained devices, participants were reoriented with the phrase, "Now, you are going to rate some more items that you rated before," and then presented with the T2 prompt for unexplained devices: "Please rate how well you feel you understand X".

Explanation Prompt Participants were instructed to write an explanation with the following paragraph (adapted from Rozenblit and Keil, 2002): "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."

Explanation Quality Participants' perceived completeness of their explanations was determined by asking them to both estimate the amount of information they generated in their explanations (% Complete) and to rate the completeness of both big, important details (Big Details) and small, less important details (Small Details) in their explanations. For exact question wording, see Wilson and Marsh (2023).

Explanation Coding Participant explanations were coded independently by two separate research assistants for the total number of causal links. Coders met to resolve any disputes and the few unresolved disagreements were settled by a third party. A causal link was defined as "the presence or inference of a part acting on another part" (Wilson & Marsh, 2023), and included three components: an acting part, the action, and the receiving part.

Look Up Questions Participants were asked if they had looked up information about any of the items they rated with the following question: "Sometimes people look up how things work because they have to fix something or because they're interested in how something works. 'Looking up' could include watching a YouTube video about how a device works, reading a website, talking to a family member or friend, talking to an expert, or any other place where you could get information about how a device works. Which of the following best fits the description of how often you have looked up information about the following items? (Please be as honest as possible. Your response will not affect your credit for this study in any way.)" Participants rated all devices with one of the following choices – Never, Once or twice before, Within the last month, Within the last week.

Screening Questions and Demographics Participants were asked two screening questions to determine how well they understood the experiment. The two questions were: "What was the current study about?" and "Please describe what you did during the study." In addition, participants were asked general demographic questions.

Procedure

All parts of this experiment were performed in-person on a lab-provided computer. Participants first consented to the study and then read the instructions and example for rating their understanding. They next rated their understanding of all twelve devices (T1 ratings). All participants were given the same set of 12 items.

Next, participants generated an explanation for one of the 12 items they just rated and then immediately re-rated their understanding of that particular item (T2 explained ratings). Participants continued to explain and then re-rate 5 additional test items. Finally, participants were asked to re-rate the final 6 items that they had previously rated but not explained, in the same one-by-one format in which they re-rated the explained items (T2 unexplained ratings).

Half of the participants were asked to explain six of the stimuli set (can opener, piano keys, toilet, zipper, spray-bottle, freezer) and simply asked to rate (without explaining) the additional six stimuli (ballpoint pen, lock and key, toaster, printer, faucet, electric blanket). The other half of participants had the explained and unexplained stimuli reversed. The item order during the T1 ratings, T2 unexplained ratings, and the order in which test items were presented to be explained and re-rated were randomized for each participant.

After completing the T2 ratings, participants were asked the three explanation quality questions with all questions for one device being asked at one time. The order of device being asked about was randomized for each participant. Finally, participants answered the look up, demographic, and screening questions.

Results

We excluded data for participants that responded to our look up question that they had looked up that device "Within the last week". Data specific to that device only was removed from analysis.[2] In total, we removed one set of T1 and T2 ratings from three different participants, across the 1656 device ratings.

IOED and Carryover

Our work showed a large amount of variability for device ratings and that simply averaging across all stimuli as often done in previous research can mask the IOED, or lack-thereof, for some stimuli (Wilson & Marsh, 2023). In order to take this variability among devices into account, we used a linear mixed modelling (LMM) approach with a device-level (can opener, toaster, etc.) random intercept. In addition, a participant-level random intercept was added to the model to account for global variability in participants' ratings, as some individuals may generally provide higher ratings than others. We focused on the ANOVA (*F*-style) results of the analyses to determine significant main effects and interactions. We followed up significant interactions with Sidak-corrected comparisons.

The first goal in these analyses was to determine whether the knowledge reassessment that occurs from writing out explanations for particular devices leads to a more generalized knowledge reassessment of unexplained devices. To this end, we ran a LMM with Time (T1 vs. T2) and Item Type (explained vs. unexplained) as factors and understanding ratings for all stimuli as the dependent measure. We found a main effect of Time, $F(1, 823) = 200.3$, $p < .001$, with T2 ratings ($M = 3.40$, $SE = 0.23$) being significantly lower than T1 ratings ($M = 4.12$, $SE = 0.23$). There was also a main effect of Item Type, $F(1,744.2) = 5.34$, $p = .021$, with average understanding ratings for explained devices ($M = 3.85$, $SE = 0.23$) being significantly higher than average understanding ratings for unexplained devices ($M = 3.67$, $SE = 0.23$).

These main effects should be interpreted in the light of a significant interaction, $F(1,823) = 14.56$, $p < .001$. Follow-up tests found that understanding ratings were significantly lower at T2 for both explained and unexplained devices (*p*s < .001). However, in comparing across item types, we found that ratings did not differ at T1 for explained and unexplained items ($p = .913$), but did differ at T2 ($p < .001$; Figure 1).

Explanation Quality and Explained Devices

The second goal for this experiment was to determine if explanation quality of explained items predicted the difference in understanding ratings for unexplained devices. First, a LMM analysis with a focus on the regression output was used to examine the relationship between explanation

[2] Inclusion of this data did not alter the significance of the results.

quality and the decrease in ratings for explained devices. A random intercept was added for both devices and subjects. Change score, calculated by subtracting T1 ratings from T2 ratings for explained devices only, was used as the dependent measure. Predictor variables included the three perceived measures of explanation completeness (% Complete, Big Details, and Small Details - all person-centered) as well as the number of causal links (which was not centered due to its natural zero point). In addition, the average of T1 and T2 ratings for explained items (grand-mean centered) was added as a covariate to account for the global variability in participants' ratings that is lost when calculating the change score. Lastly, person-level means for participants' completeness ratings were added as covariates to the model to account for the person-centering of the three completeness ratings (Field, 2018). When the initial model was run, the device random intercept was determined not to significantly improve the model (Wald statistic, z = 1.11, p = .267), therefore, it was removed from the model.

The results of the LMM are shown in Table 1 for the predictors of relevance under "Four Predictor Model". Two of the four predictors were found to be significant - % Complete and Causal Links. The model was run again including only significant predictors (as well as the person-level mean for % Complete), with the rest of the model kept the same. Both % Complete and Causal Links were, again, found to be significant in the "Final Model" (Table 1). In short, participants who gave lower ratings for % Complete and participants who had less causal links in their explanations both saw a larger decrease in understanding ratings from T1 to T2 for explained devices.

Explanation Quality and Unexplained Devices

We next tested whether any of the explanation quality variables predicted the observed decrease in understanding ratings for unexplained devices. For this analysis, a LMM was constructed comparing predictor variables to the difference in understanding ratings for unexplained devices.

Change score, calculated by subtracting T1 ratings from T2 ratings for unexplained devices only, was the dependent measure. Predictor variables for this model were calculated by taking the average of the variable for each participant for all explained items (these items were grand-mean centered).

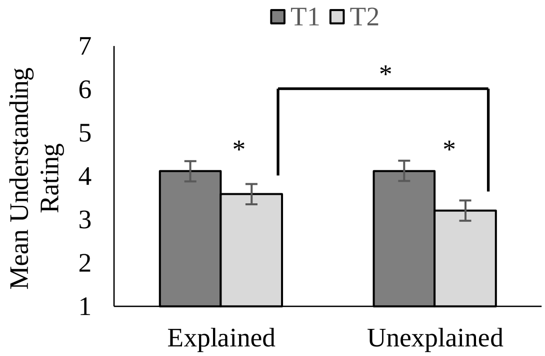

Figure 1: The average understanding ratings at T1 and T2 for explained and unexplained items. Error bars indicate standard error. * $p < .001$

A random intercept for devices and subjects, and the average of T1 and T2 ratings for unexplained devices per participant (grand-mean centered) were added to the model. Importantly, all data pertaining to explanation quality was taken from explained devices while all data pertaining to understanding ratings was taken from unexplained devices.

The model was unable to converge with both random intercepts, so the random intercept for devices was removed from the model. The results of the final LMM are shown in Table 2 for the four predictor variables – none of which were found to have a significant relationship with the change in understanding ratings for unexplained devices.

General Discussion

While the exposure of the IOED through generating a causal explanation has been repeatedly shown (e.g., Fernbach et al., 2013; Lawson, 2006; Zeveney & Marsh, 2016), there has been little investigation into the effect of this knowledge reassessment on other, unexplained concepts and how explanation quality may affect such knowledge transfer. We found that the knowledge reassessment that occurs through the act of writing a causal explanation is transferable to the causal knowledge of other, unexplained items, which replicates the results found in Meyers et al. (2023). Interestingly, our results show that, although no difference is seen in initial ratings for explained versus unexplained

Table 1: Predictive ability of explanation quality for explained items.

Predictor	Estimate	SE	95% CI Lower	95% CI Upper	df	t	p
Four Predictor Model							
% Complete	0.026	0.006	0.014	0.038	369.5	4.26	**< .001**
Big Details	0.123	0.070	-0.015	0.260	343.8	1.76	.080
Small Details	0.108	0.070	-0.029	0.245	341.7	1.56	.120
Causal Links	0.073	0.021	0.032	0.113	400.8	3.50	**< .001**
Final Model							
% Complete	0.036	0.004	0.027	0.044	408.0	8.23	**< .001**
Causal Links	0.079	0.021	0.039	0.120	406.2	3.81	**< .001**

Table 2: Predictive ability of explanation quality for unexplained items.

Predictor	Estimate	SE	95% CI		df	*t*	*p*
			Lower	Upper			
Avg % Complete	0.018	0.010	-0.002	0.039	63.73	1.81	.075
Avg Big Details	-0.035	0.113	-0.260	0.190	62.76	-0.31	.758
Avg Small Details	0.148	0.132	-0.115	0.411	62.93	1.12	.265
Avg Causal Links	0.016	0.055	-0.094	0.126	63.03	0.29	.772

devices, there was a greater decrease in understanding ratings for unexplained items than explained items. This adds further clarification that the decrease seen for unexplained devices is actually larger than explained devices.

Overgeneralization?

So why is there a significantly greater drop for the unexplained items? One could have predicted less of a drop if unexplained items in an IOED paradigm are processed similarly to the reflection group in Johnson et al. (2016, Experiments 1 and 5), where participants were asked to simply reflect on their explanatory ability for particular devices. However, there is a key difference in the experience of participants in our experiment (within-subjects) versus Johnson et al.'s (2016) study (between-subjects) - our participants initially experienced failure in relation to providing explanations. That is, participants generated an explanation that they knew to be low in quality for the explained devices. It is possible that this initial failure to perform a task worked to further decrease participants' confidence when asked about similar devices.

Interestingly, explanation quality variables found to be predictive of the drop in ratings for explained devices were not found to have a relationship with the decrease in ratings also seen for unexplained devices. This could be due to study limitations such as the simplicity of the analysis performed (i.e., comparing the averages of the variables across devices), and future studies with a larger sample size may be able to perform more specified analyses that account for the lost variability.

Future Directions

Our research, along with its limitations, suggest several future avenues of exploration. First, is this knowledge reassessment retained over time? Future work in our lab addresses this question by having participants return one week post-IOED paradigm to make an additional Time 3 (T3) understanding rating. Comparing T1, T2, and T3 values provides insight into whether the knowledge reassessment induced by the IOED paradigm persists or the original illusion of knowledge returns.

One limitation of our experiment is that we intermixed simple (e.g., can opener) and complex (e.g., printer) devices. As such, we were not able to evaluate how the complexity of an explained device may influence perception of other devices. Johnson et al. (2016, Experiment 5) looked at changes in understanding ratings for simple (Velcro, reading glasses) versus complex (vacuum cleaner, computer mouse) devices in both an explanation generation condition and a reflection condition. They found that device complexity had little effect on knowledge reassessment for the explanation condition, but mere reflection on the inner workings of complex devices caused a significantly greater reduction in understanding ratings than for simple devices. These findings would suggest that we may not see any difference by item complexity in our experiment for explained items, but the influence of failing to explain simple versus complex devices may influence transfer to unexplained items. For instance, if the person had failed to explain how a manual can opener works, they may think that they surely have no shot at explaining a particle accelerator. In ongoing work in our lab, we are exploring item complexity to determine if any shock to the system of lacking understanding may function to recalibrate causal knowledge more generally, or if it depends on the item.

Additionally, we only explored transfer in the one domain (devices). Meyers et al.'s (2023) final experiment looked at knowledge transfer between devices and natural phenomena and found results suggestive of transfer across domains. Future work can test the magnitude of this carryover across similar and different domains. For example, would failing to explain how a can opener works suggest to a person that they do not understand how a mental disorder develops (see Zeveney & Marsh, 2016)?

Conclusion

Ignorance is not always blissful. When considering our awareness of how common household devices work, as in this study, the IOED may seem relatively harmless. However, ignorance to our lack of knowledge can lead us to make miscalibrated and ill-informed life-altering decisions (Alter et al., 2010), such as which politician to vote for, whether to receive medical treatment, or to seek / decline expert advice (Scharrer et al., 2014) – decisions that, later, may not have been in our best interest.

We have replicated previous findings that when someone learns of their causal knowledge failings, it can spread to other items. Our results suggest this transfer of knowledge may even be overgeneralized. In this way, being shown their lack of knowledge in one area may help people concede their lack of knowledge in others. Future studies are needed to further flesh-out the boundaries of metacognitive monitoring on knowledge reassessment, but these results show promise toward freeing ourselves from the ignorance of our ignorance.

References

Alter, A. L., Oppenheimer, D. M., & Zemla, J. C. (2010). Missing the trees for the forest: A construal level account of the illusion of explanatory depth. *Journal of Personality and Social Psychology*, *99*(3), 436–451. https://doi.org/10.1037/a0020218

Bes, B., Sloman, S., Lucas, C. G., & Raufaste, É. (2012). Non-Bayesian inference: Causal structure trumps correlation. *Cognitive Science*, *36*(7), 1178–1203. https://doi.org/10.1111/j.1551-6709.2012.01262.x

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51–64. https://doi.org/10.1037/xge0000505

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

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications Limited.

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

Johnson, D., Murphy, M., & Messer, R. (2016). Reflecting on explanatory ability: A mechanism for detecting gaps in causal knowledge. *Journal of Experimental Psychology: General*, *145*(5), 573–588. https://doi.org/10.1037/xge0000161.supp

Koriat, A., & Levy-Sadot, R. (2000). Conscious and unconscious metacognition: A rejoinder. *Consciousness and Cognition*, *9*, 193–202. https://doi.org/10.1006/ccog.2000.0436

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

Meyers, E. A., Gretton, J. D., Budge, J. R. C., Fugelsang, J. A., & Koehler, D. J. (2023). Broad effects of shallow understanding: Explaining an unrelated phenomenon exposes the illusion of explanatory depth. *Judgment and Decision Making*, *18*(24), 1-18. https://doi.org/10.1017/jdm.2023.24

Mills, C. M. & Keil, F. C. (2004). Knowing the limits of one's own understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, *87*(1), 1-32. https://doi.org/10.1016/j.jecp.2003.09.003

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, *46*(2), 168–175. https://doi.org/10.1177/0098628319834381

Roeder, S. (2016). The disparity between what we know and how we communicate. [Doctoral dissertation, University of California, Berkeley] UC Berkeley Campus eScholarship. https://escholarship.org/uc/item/35j6b5dz

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

Scharrer, L., Stadtler, M., & Bromme, R. (2014). You'd better ask an expert: Mitigating the comprehensibility effect on laypeople's decisions about science-based knowledge Claims. *Applied Cognitive Psychology*, *28*(4), 465–471. https://doi.org/10.1002/acp.3018

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00218

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

Wilson, J., & Marsh, J. K. (2023). Perceptions of explanation completeness help decrease knowledge overestimation. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Meeting of the Cognitive Science Society* (pp. 717–723). Cognitive Science Society. https://escholarship.org/uc/item/9gv611vd

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* (pp. 1020-1025). Cognitive Science Society.