



Cognitive Science (2015) 1–19

Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12280

The Curse of Expertise: When More Knowledge Leads to Miscalibrated Explanatory Insight

Matthew Fisher, Frank C. Keil

Department of Psychology, Yale University

Received 10 March 2015; received in revised form 19 May 2015; accepted 23 May 2015

Abstract

Does expertise within a domain of knowledge predict accurate self-assessment of the ability to explain topics in that domain? We find that expertise increases confidence in the ability to explain a wide variety of phenomena. However, this confidence is unwarranted; after actually offering full explanations, people are surprised by the limitations in their understanding. For passive expertise (familiar topics), miscalibration is moderated by education; those with more education are accurate in their self-assessments (Experiment 1). But when those with more education consider topics related to their area of concentrated study (college major), they also display an illusion of understanding (Experiment 2). This “curse of expertise” is explained by a failure to recognize the amount of detailed information that had been forgotten (Experiment 3). While expertise can sometimes lead to accurate self-knowledge, it can also create illusions of competence.

Keywords: Expertise; Ignorance; Explanation

1. Introduction

Do expertise and learning invariably cause greater insight into one’s own abilities and limitations? Such a view is frequently used to promote the benefits of higher education. Those who are more informed about a topic are thought to be less dogmatic because they better realize the limits of their own understanding; they move toward Socrates’ form of true wisdom: “knowing you know nothing.” Yet we have all encountered individuals with decades of immersion in an area of expertise who nevertheless possess unrealistically inflated senses of their own knowledge. Here, we argue that both insight and illusion into one’s explanatory competence can co-exist and that they occur in systematic ways related to the kind of expertise involved.

Correspondence should be sent to Matthew Fisher, Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520-8205. E-mail: matthew.fisher@yale.edu

Explanatory understanding is a critical component of cognition; it undergirds many processes such as categorization, diagnosis, and induction (Gopnik & Wellman, 1994; Keil, 1998, 2006; Murphy, 2000; Murphy & Medin, 1985; Williams & Lombrozo, 2013). However, we often overestimate our explanatory prowess, exhibiting an illusion of explanatory depth (IOED). In the IOED paradigm, participants initially rate their explanatory ability (Time 1) and then after writing out as complete of an explanation as they can, they re-rated their ability (Time 2). People show a consistent drop from Time 1 to Time 2 in their reported understanding of such things as common artifacts (Rozenblit & Keil, 2002; Wilson & Keil, 1998), word meanings (Kominsky & Keil, 2014), and political issues (Fernbach, Rogers, Fox, & Sloman, 2013; Fisher & Keil, 2014). The IOED is one facet of a broader family of phenomena in which people make inaccurate self-assessments, often being far more confident about their abilities than is warranted (Dunning, 2005; Koriat & Bjork, 2005).

The relationship between expertise and overconfidence is not straightforward. It sometimes is associated with reductions in overconfidence and other times with increases (for reviews, see Camerer & Johnson, 1997; Koehler, Brenner, & Griffin, 2002). These mixed results may be better understood by considering two different types of expertise. First, *passive expertise*: a type of expertise arising from exposure through life experience and the position one occupies in a society or culture. Age or gender, for example, can influence the sorts of passive expertise one acquires. Second, *formal expertise*: a type of expertise arising from extended study of a particular topic. Formal expertise is a self-selected domain of knowledge, deliberately mastered, with clear benchmarks of success. This type of expertise can be thought of as the type of knowledge acquired after completing a college major. One key feature of formal expertise (explored in Experiment 3) is that after a period of dedicated study, many of the details acquired during extended study are forgotten over time.

This distinction between types of expertise may illuminate some reasons behind the illusions as well as suggest ways to mitigate them in cases where they might have serious negative consequences. The illusions may persist for multiple reasons (Alter, Oppenheimer, & Zemla, 2010; Bjork, 1999; Kelley & Lindsay, 1993), but here we investigate how different types of expertise influence self-evaluations of explanatory insight. This approach builds on prior findings of inaccurate self-evaluations in explicitly studied domains (Mabe & West, 1982). For example, trained nurses overestimate their ability to perform basic life support skills (Wynne, Marteau, Johnston, Whiteley, Evans, 1987) and general practitioners do not correctly assess their own medical knowledge (Tracey, Arroll, Richmond, & Barham, 1997). Students inaccurately judge their own progress; over the course of medical school, the correlations weaken between students' self-evaluations and faculty supervision ratings (Arnold, Willoughby, & Calkins, 1985). Those who claim to have some expertise are often more confident and less willing to admit they do not know (Bradley, 1981) and self-classification or "self-views" as an expert leads to poor accuracy in metacognitive judgments (Critcher & Dunning, 2009; Dunning, Johnson, Ehrlinger, & Kruger, 2003; Glenberg & Epstein, 1987).

These findings, however, create a tension with another phenomenon: Those who are the least competent in a domain are typically also those who are the least aware of their

incompetence (Dunning, 2011; Kruger & Dunning, 1999; Maki, Jonas, & Kallod, 1994; Miller & Geraci, 2011). Thus, general incompetence, lack of ability, or less overall education (or all three) may be related to especially large disconnects between what one thinks one knows and what one really knows. Yet how could greater expertise lead to overconfidence if overconfidence is bred by ignorance? Considering distinctive types of expertise may help provide an answer.

Formal expertise may lead to relatively accurate self-assessment for a wide variety of topics because it enables people to be more aware in general of how much knowledge is required to fully understand even very familiar topics. We therefore propose that those with formal expertise are less likely to exhibit overconfidence for everyday explanations than those with passive expertise only. Formal expertise, however, may exacerbate the IOED for topics within a chosen domain of expertise. Areas of concentrated study (i.e., college major) differ in that people deliberately focus on understanding explanatory detail as they acquire the formal expertise. This previous mastery, which can quickly be forgotten, may later give rise to illusions of understanding because of a failure to accurately understand the rate of forgetting. Thus, expertise may be related to greater insight about one's knowledge limitations in general but "curse" expert assessments in one's own areas of expertise. When considering the influence of knowledge on illusions of understanding, we therefore predict:

Hypothesis 1: People are quite accurate regardless of education at assessing their knowledge for unfamiliar areas of passive expertise, but only those with lower levels of education overestimate their knowledge of familiar areas.

Hypothesis 2: Highly educated individuals tend to overestimate their ability to explain their own areas of formal expertise more than they do unfamiliar areas of formal expertise or familiar areas of passive expertise.

To examine these issues, we explored two forms of expertise: passive expertise arising from exposure because of the position one occupies in a culture and which is largely not chosen (such as one's age or gender; Experiment 1) and formal expertise arising from extended study of a particular topic (Experiment 2). While perhaps reducing illusions for some topics, formal expertise may also have the effect of increasing the illusion of retaining explanatory details in one's specific area of concentrated study. We present evidence that this "curse of expertise" is driven by a failure to realize how quickly explanatory information has been forgotten (Experiment 3).

2. Experiment 1

2.1. *Methods*

To test these hypotheses, we adopted the task from previous IOED research (Rozenblit & Keil, 2002), using gender-based and age-based topics as proxies for passive expertise. These topics were ones to which people had been exposed through everyday experiences

but which they had not spent time explicitly studying the explanatory details. According to our predictions, in these areas of passive expertise, those without formal expertise (college education), should show the IOED for familiar topics, while those with formal expertise should be well calibrated.

2.1.1. Participants

First, participants were divided into three age groups: 18–25, 26–49, and 50+. We chose to test two areas of passive expertise, age and gender, for which we expected to find similar results. Older and younger adults were used to test the effect of knowledge on age-related topics and the middle range of ages was divided by gender to test the effect of knowledge for gender-related topics. See Table 1 for detailed demographic information. Based on pilot data, it was determined that 50–90 participants per condition would be needed to detect an effect.

All participants completed the study online through Amazon's Mechanical Turk (Rand, 2012) and lived in the United States.

2.1.2. Design

Experiment 1 used a Time (within-subjects) \times Knowledge \times Education \times Domain mixed model design. After providing Time 1 ratings for how well they could explain topics in a domain of passive expertise (age-based or gender-based topics), participants were randomly assigned to fully explain either three familiar topics (High Knowledge) or

Table 1
Demographic information for participants in Experiment 1

	Age (years)		
	18–25	26–49	50+
<i>Male</i>			
Bachelor's degree			
<i>N</i>	29	42	10
<i>M</i> _{age}	23.24	31.64	64
<i>SD</i> _{age}	1.5	5.03	8.14
No bachelor's degree			
<i>N</i>	45	39	10
<i>M</i> _{age}	21.82	32.26	57.4
<i>SD</i> _{age}	1.86	5.88	4.77
<i>Female</i>			
Bachelor's degree			
<i>N</i>	22	34	15
<i>M</i> _{age}	23.5	32.74	57
<i>SD</i> _{age}	1.1	6.33	4.91
No bachelor's degree			
<i>N</i>	33	42	15
<i>M</i> _{age}	21.85	33.93	58.2
<i>SD</i> _{age}	1.95	6.01	8.26

three less familiar topics (Low Knowledge). The High Knowledge topics related to their own group, be it older or younger, or male or female, and the Low Knowledge topics related to the other group (i.e., a female explaining male topics). After writing out an explanation, all participants re-rated their ability to explain that topic (Time 2) using the same scale they had used at Time 1. The analysis took into account participants' education level, separating the results from the More Education participants from the Less Education participants.

2.1.3. Procedure

In order to familiarize participants with the scale used to judge explanations, participants first went through a short training procedure adopted from the original IOED paradigm (Rozenblit & Keil, 2002). Participants read two sets of examples of high, middle, and low quality explanations (see Appendix S1 in Supplementary Materials) and rated them on a 1 (Very poor) to 7 (Very good) Likert scale. For the first set of explanations, if the participants rated any of the example explanations incorrectly, an alert appeared asking them to reassess the explanation. They could not continue until they had properly scored each explanation. They did not receive feedback for the second set of explanations, but their scores were recorded to ensure they had learned how to properly use the scale. The exclusion of the small percentage of participants who gave inaccurate responses to the training items did not change the result of any experiment, so the results of all participants are reported.

To test how well people can explain areas of passive expertise, we compiled a list of topics that would be familiar to participants based on their position in society (age or gender), not their deliberately chosen areas of formal education. In a separate pre-test, we identified 10 topics people thought men could explain better than women (e.g., plumbing, cars), 10 topics people thought women could explain better than men (e.g., pregnancy, cooking), 10 topics people thought older adults could explain better than younger adults (e.g., retirement, cassette players), and 10 topics people thought younger adults could explain better than older adults (e.g., smart phones, the Internet). See Appendix S2 in Supplementary Materials for the full lists of topics. For the Time 1 ratings, participants either received 20 gender-related or age-related topics to consider. Half of the items in each set of 20 were High Knowledge topics (a female participant considering female topics) and half were Low Knowledge topics (a female participant considering male topics). At Time 1, they were asked to "consider your ability to explain [topic]. If you were to write an explanation to a group of laypeople (e.g., other Amazon Turk workers) who have been given the same training to use the scale, what rating would your explanation be given?". Participants responded on a scale from 1 (Very poor) to 7 (Very good) for the 20 topics. They were instructed to not pause excessively on the items to prevent participants from fully generating their responses to themselves before responding.

After these initial ratings (Time 1), participants wrote out explanations for a random set of either three of the High Knowledge topics or three of the Low Knowledge topics they had rated at Time 1. Thus, they were randomly assigned to the High Knowledge (e.g., a female writing an explanation for female topics) or the Low Knowledge condition

(e.g., a female writing an explanation for male topics). After each written explanation, participants rerated their ability to explain the topics they had just explained (Time 2) using the exact scale used at Time 1.

2.2. Results

A mixed model ANOVA examined the effect of Time (Time 1, Time 2), Knowledge (High Knowledge, Low Knowledge), and Education (More Education, Less Education)¹ on participants' explanation self-assessments (see Table 2). Because there were no meaningful differences based on the domain of passive expertise (Age vs. Gender), both types of passive expertise were collapsed together for the purpose of the analysis. There was a marginally significant Time \times Education \times Knowledge interaction, $F(1, 332) = 2.86$, $p = .09$. This was driven by a Time \times Knowledge interaction for Less Education participants, $F(1, 182) = 7.58$, $p = .006$, but not for the More Education participants, $F(1, 150) = .22$, $p = .64$. Thus, the effect of passive expertise on illusory understanding was primarily driven by the Less Education participants in the High Knowledge condition, who showed a difference between their Time 1 ($M = 3.72$, $SD = 1.45$, 95% CI = [3.41, 4.03]) and Time 2 ($M = 3.10$, $SD = 1.39$, 95% CI = [2.80, 3.40]) ratings, $t(81) = 4.35$, $p < .001$, Cohen's $d = 0.44$, but were well calibrated in the Low Knowledge condition—showing no significant drop from Time 1 ($M = 2.73$, $SD = 1.27$, 95% CI = [2.48, 2.98]) to Time 2 ($M = 2.58$, $SD = 1.26$, 95% CI = [2.34, 2.83]), $t(101) = .14$, Cohen's $d = 0.12$ (see Fig. 1). This provides partial support for Hypothesis 1: We found a trending effect whereby passive expertise leads to illusions of understanding for those with low education, but not those with formal expertise.

3. Experiment 2a

3.1. Methods

Although those with more education did not show an illusion of understanding for familiar topics, they may do so for areas of formal expertise, domains of knowledge that

Table 2
Mean and *SD* of explanation ratings by education level

Education Level	High Knowledge		Low Knowledge	
	Mean	<i>SD</i>	Mean	<i>SD</i>
More Education				
Time 1	3.63	1.39	2.63	1.39
Time 2	3.48	1.48	2.55	1.32
Less Education				
Time 1	3.72	1.45	2.73	1.27
Time 2	3.1	1.39	2.58	1.26

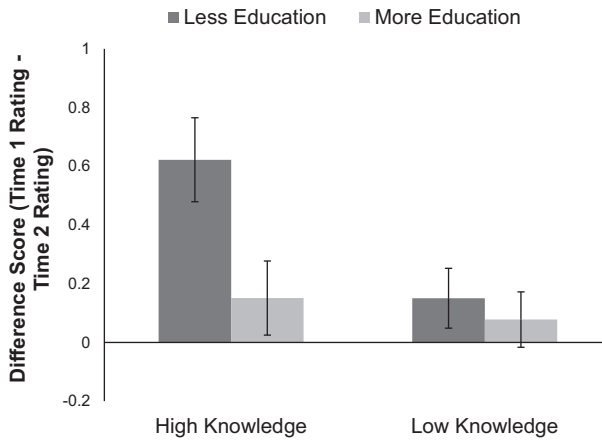


Fig. 1. Difference scores for explanation ratings as a function of education level and knowledge. Error bars, mean \pm 95% CI.

have been intentionally learned in detail. Experiment 2 used topics of formal expertise (college major)—domains in which college-educated people have intentional intensive exposure rather than passive experience through their position in life (e.g., gender or age). We predicted that chosen area of expertise, indicated by college major, would lead to miscalibrated knowledge for that particular domain.

3.1.1. Participants

One hundred and fifteen participants (61 males, 54 females; $M_{\text{age}} = 31.54$, $SD = 10.16$) completed Experiment 2a. All participants completed the study online through Amazon's Mechanical Turk and lived in the United States. Only participants who had completed a bachelor's degree were eligible for Experiment 2. Participants did not complete multiple experiments; each experiment contains a unique naïve sample.

3.1.2. Procedure

Experiment 2 used the same procedure as the previous experiment with minor changes. At Time 1, participants rated their ability to explain three topics from the 10 most popular college majors, making a total of 30 topics (e.g., Krebs's cycle, thermodynamics, object-oriented programming). After making Time 1 ratings, participants were randomly assigned to either write out full explanations for the three topics within their major or a random subset of three topics outside of their major. After writing each explanation, participants rerated their ability to explain that topic using the same scale from Time 1.

3.2. Results

A mixed model ANOVA examined the effect of Time (Time 1, Time 2) and Knowledge (High Knowledge, Low Knowledge) on participants' explanation ratings. As expected,

explanations from the High Knowledge condition received higher ratings ($M = 4.44$, $SD = 1.26$, 95% CI = [4.21, 4.67]) than the explanations from the Low Knowledge condition ($M = 2.57$, $SD = 1.13$, 95% CI = [2.36, 2.78]), $F(1, 113) = 67.50$, $p < .001$, Cohen's $d = 1.56$. As we found with the less educated participants from Experiment 1, there was a Time \times Knowledge interaction $F(1, 113) = 4.08$, $p = .05$, driven by a larger difference between Time 1 ($M = 4.93$, $SD = 1.37$, 95% CI = [4.60, 5.26]) and Time 2 ($M = 3.97$, $SD = 1.47$, 95% CI = [3.62, 4.32]) ratings in the High Knowledge condition $t(66) = 6.01$, $p < .001$, Cohen's $d = 0.68$, providing support for Hypothesis 2. Participants in the Low Knowledge condition also dropped their ratings from Time 1 ($M = 2.83$, $SD = 1.35$, 95% CI = [2.44, 3.22]) to Time 2 ($M = 2.31$, $SD = 1.09$, 95% CI = [1.93, 2.69]), $t(47) = 3.67$, $p < .001$, Cohen's $d = 0.43$, but not to the same extent as those in the High Knowledge condition (see Table 3 and Fig. 2). These results show miscalibrated assessments of one's knowledge within a domain of expertise. Participants realized they could not produce the quality of explanation they initially thought they could for topics within their area of study.

Table 3
Mean and SD of explanation ratings by amount of knowledge

	Mean	SD
High Knowledge		
Time 1	4.93	1.37
Time 2	3.97	1.47
Low Knowledge		
Time 1	2.83	1.35
Time 2	2.31	1.09

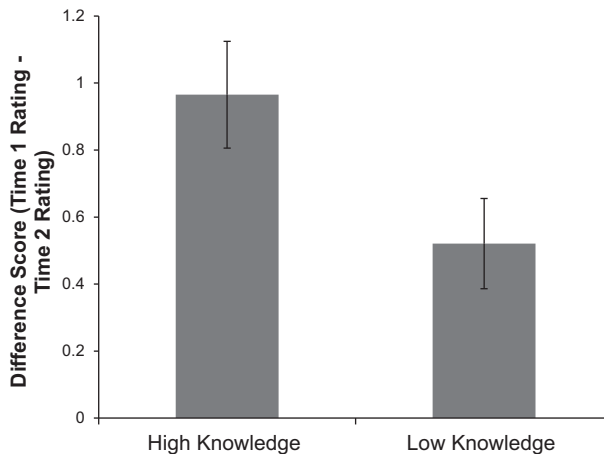


Fig. 2. Difference scores for explanation ratings as a function of knowledge. Error bars, mean \pm 95% CI.

4. Experiment 2b

4.1. Methods

We next directly compared the assessment of passive and formal expertise within a population of “experts” (college graduates). Experiment 2b replicated the results of Experiment 1 and 2a and provided a strong test of Hypothesis 2. In Experiment 2b, participants either rated their ability to explain topics from within their college major (formal expertise) or they rated their ability to explain unfamiliar everyday topics (passive expertise).

4.1.1. Participants

Forty-five participants (24 males, 21 females; $M_{\text{age}} = 32.91$, $SD = 10.99$) were assigned to the Formal Expertise condition in Experiment 2b. Fifty-four participants were assigned to the Passive Expertise Condition (see Table 4 for demographic information). All participants completed the study online through Amazon’s Mechanical Turk and lived in the United States. Only participants who had completed a bachelor’s degree were eligible for Experiment 2b.

4.1.2. Procedure

Experiment 2b combined the procedures from Experiments 1 and 2a. First, participants were randomly assigned to Formal Expertise condition or the Passive Expertise condition. Participants in the Formal Expertise condition provided Time 1 ratings for topics from 10 college majors, wrote out full explanations for the three topics within *their* major, and then rerated their ability to explain those three topics (as in Experiment 2a). Participants in the Passive Expertise condition ages 18–25 and over 50 received the age-related topics, while the rest received gender-related topics. They provided Time 1 ratings for either all age-related or gender-related topics, wrote out full explanations for a random subset of three topics inside their domain of passive expertise, and then rerated their ability to explain those three topics (as in Experiment 1).

Table 4
Demographic information for participants in the passive expertise condition of Experiment 2b

	Age (years)		
	18–25	26–49	50+
Male			
<i>N</i>	2	20	4
<i>M</i> _{age}	24	34.55	63.75
<i>SD</i> _{age}	1.41	7.01	4.27
Female			
<i>N</i>	4	18	6
<i>M</i> _{age}	23	32.26	56
<i>SD</i> _{age}	1.63	5.88	3.22

Table 5
Mean and *SD* of explanation ratings by type of expertise

	Mean	<i>SD</i>
Formal expertise		
Time 1	5.17	1.08
Time 2	4.14	1.2
Passive expertise		
Time 1	3.49	1.25
Time 2	3.49	1.2

4.2. Results

A mixed model ANOVA examined the effect of Time (Time 1, Time 2) and Expertise (Formal Expertise, Passive Expertise) on participants' explanation ratings. Like Experiment 1, age-related and gender-related topics were collapsed together in the Passive condition. Formal Expertise explanations received higher ratings ($M = 4.66$, $SD = 1.25$, 95% CI = [4.41, 4.91]) than Passive Expertise explanations ($M = 3.49$, $SD = 1.22$, 95% CI = [3.25, 3.73]), $F(1, 99) = 29.21$, $p < .001$, Cohen's $d = 0.95$. Again, we found a Time \times Expertise interaction $F(1, 99) = 26.37$, $p < .001$ driven by a larger difference between Time 1 ($M = 5.17$, $SD = 1.08$, 95% CI = [4.86, 5.48]) and Time 2 ratings ($M = 4.14$, $SD = 1.20$, 95% CI = [3.79, 4.49]) in the Formal Expertise condition, $t(45) = 6.76$, $p < .001$, Cohen's $d = 0.90$, and no difference between Time 1 ($M = 3.49$, $SD = 1.25$, 95% CI = [3.16, 3.83]) and Time 2 ($M = 3.49$, $SD = 1.20$, 95% CI = [3.16, 3.81]) ratings in the Passive Expertise condition $t(54) = .05$, $p = .96$, Cohen's $d = 0.0$ (see Table 5 and Fig. 3). This finding replicated the results of the previous experiments and provided strong evidence in favor of Hypothesis 2: Formal expertise leads to illusions of understanding for previously mastered formal topics.

5. Experiment 3a

We propose that a failure to recognize the rapid rate of memory decay for a particular topic gives rise to the "curse of expertise." When educated participants in Experiment 2 inaccurately predicted how well they could explain phenomena from their area of study, it may be due to "meta-forgetfulness": a failure to realize how much has been forgotten since they had maximum mastery of a topic. They may only sense a gradual tapering of explanatory details when in fact the drop-off after mastery occurs quickly and is very substantial. Self-assessments of memory are only weakly correlated with actual performance; thus, people fail to recognize the decay of their own memories (Herrmann, 1982). Furthermore, although memory declines with age, the elderly report that forgetfulness does not impact them at all (Sunderland, Watts, Baddeley, & Harris, 1986). The act of retrieval gives rise to the illusion of truth, where the fluency of the memory is mistaken

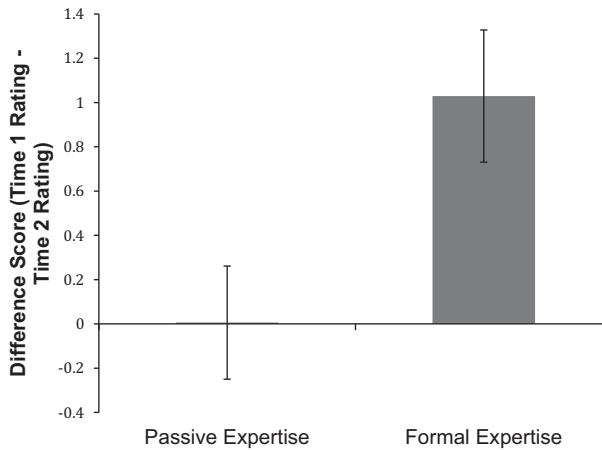


Fig. 3. Explanation ratings for different types of expertise. Error bars, mean \pm 95% CI.

for validity (Ozubko & Fugelsang, 2011). Similarly, in our task, participants may assess their explanatory ability by responding as if they had barely forgotten any of the details they once knew. Thus, the mechanism explaining the effect found in Experiment 2 could be a misattribution of the maximum amount of knowledge at some point in the past for current knowledge. To test this idea, we asked participants to also rate the quality of explanation they could provide at the peak of the knowledge for topics within their college major. If people indeed fail to account for the decay in their memory, we would expect to see no difference between ratings of “peak” knowledge and their Time 1 ratings.

5.1. Methods

5.1.1. Participants

Seventy-one participants (44 males, 27 females; $M_{\text{age}} = 32.49$, $SD = 11.16$) completed Experiment 3a. All participants completed the study online through Amazon’s Mechanical Turk and lived in the United States. Only participants who had completed a bachelor’s degree were eligible for Experiment 3a.

5.1.2. Procedure

Experiment 3a was identical to Experiment 2a except participants only considered topics within their college major and self-assessed their ability to explain a third time. Thus, participants rated their ability to explain 30 topics from college majors (Time 1), wrote out full explanations for the three topics within their own major, and rerated their ability to explain those three topics (Time 2). The following instructions appeared after each Time 2 rating, “Now consider the point in your life when you knew the most about this topic, how would you answer the following question:” Participants then reported the rating other mTurk workers would give their explanation (using the same wording as the

Time 1 and Time 2 questions used in Experiments 1 and 2). We call this rating the “Time 0” rating because it is an assessment of how well participants originally knew the explanation.

5.2. Results

Replicating the results of Experiment 2, we found that college-educated adults experienced an illusion of explanatory depth when considering topics from their college major. There was again a significant drop from their Time 1 rating ($M = 4.94$, $SD = 1.67$, 95% CI = [4.55, 5.33]) to their Time 2 rating ($M = 3.39$, $SD = 1.42$, 95% CI = [3.06, 3.72]), $t(70) = 7.29$, $p < .001$, Cohen’s $d = 1.0$. We found no difference between Time 1 ($M = 4.94$, $SD = 1.67$, 95% CI = [4.55, 5.33]) and Time 0 ratings ($M = 5.16$, $SD = 1.42$, 95% CI = [4.83, 5.49]), $t(70) = -1.36$, $p = .18$, Cohen’s $d = 0.14$, suggesting participants at Time 1 think they can give the same quality of explanation as the point in their life when they knew the most about the topic (see Fig. 4). This provides support for “meta-forgetfulness” as the mechanism giving rise to the curse of expertise: Participants fail to recognize the decay from “peak” to current knowledge.

6. Experiment 3b

6.1. Methods

Experiment 3a provided initial evidence for the underlying mechanism explaining the results of Experiment 2. However, participants may have provided high Time 0 ratings as a way to compensate for their self-admittedly sparse explanatory knowledge. If so, the correspondence between Time 1 and Time 0 would not indicate a neglect of mem-

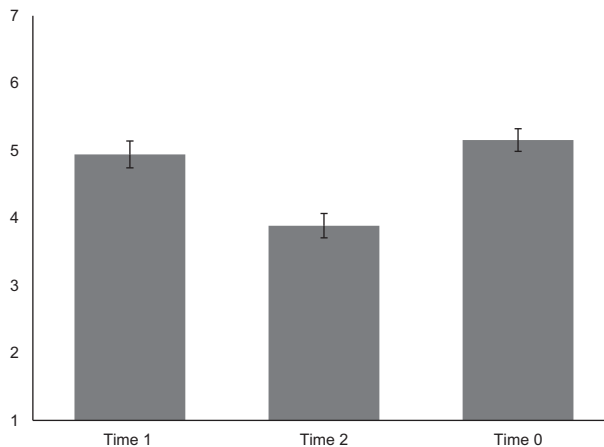


Fig. 4. Self-assessments of explanatory knowledge in Experiment 3a. Error bars, mean \pm 95% CI.

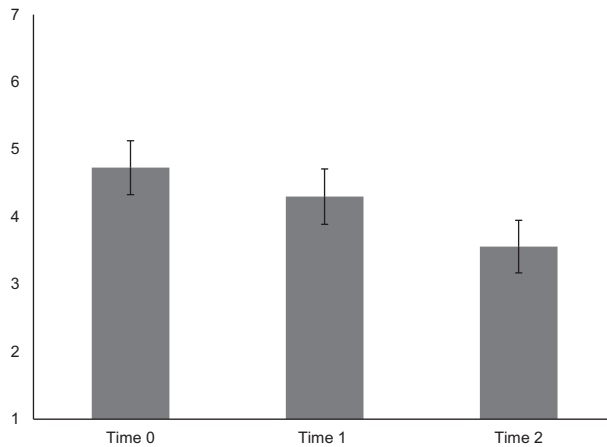


Fig. 5. Self-assessments of explanatory knowledge in Experiment 3b. Error bars, mean \pm 95% CI.

ory decay, but a way for participants to signal that even though they were unable to produce a quality explanation they could have done so at some point in the past. To help rule out this possibility, participants in Experiment 3b reported their Time 0 rating *before* Time 1. Since this was the first set of ratings they provided, there was no reason for them to alter their Time 0 results due to any other ratings or explanations in the experiment.

6.1.1. Participants

Seventy-eight participants (36 males, 42 females; $M_{\text{age}} = 31.26$, $SD = 9.86$) completed the main task in Experiment 3b. All participants completed the study online through Amazon's Mechanical Turk and lived in the United States. Only participants who had completed a bachelor's degree were eligible for Experiment 3b.

6.1.2. Procedure

Experiment 3b was identical to Experiment 3a except that participants provided Time 0 ratings before the Time 1 and Time 2 ratings.

6.2. Results

Again replicating the results of Experiment 2, participants dropped their self-assessment ratings after actually writing out full explanations. There was again a significant drop from their Time 1 rating ($M = 4.30$, $SD = 1.87$, 95% CI = [3.89, 4.71]) to their Time 2 rating ($M = 3.56$, $SD = 1.76$, 95% CI = [3.17, 3.95]), $t(77) = 7.29$, $p < .001$, Cohen's $d = 0.41$. We did find a difference between Time 0 ($M = 4.73$, $SD = 1.78$, 95% CI = [4.33, 5.13]) and Time 1 ratings ($M = 4.30$, $SD = 1.87$, 95% CI = [3.89, 4.71]), $t(77) = 3.80$, $p < .001$, Cohen's $d = 0.24$, because Time 1 ratings lowered after first considering "peak" knowledge. Due to this order effect, the Time 1 ratings from Experiment

Table 6
Mean and *SD* of explanation ratings in Experiments 3a and 3b

	Mean	<i>SD</i>
Experiment 3a		
Time 1 (unbiased)	4.94	1.67
Time 2	3.39	1.42
Time 0	5.16	1.42
Experiment 3b		
Time 0 (unbiased)	4.73	1.78
Time 1	4.3	1.87
Time 2	3.56	1.76

3a may be a better estimate of participants' unbiased response. The Time 0 ratings of Experiment 3a ($M = 4.73$, $SD = 1.78$, 95% CI = [4.33, 5.13]) were no different than the Time 1 ratings of Experiment 3a, when participants had not yet provided Time 0 ratings, ($M = 4.94$, $SD = 1.67$, 95% CI = [4.55, 5.33]), $t(148) = .77$, $p = .44$, Cohen's $d = 0.12$ (see Table 6).

6.3. Discussion

Taken together, this evidence helps interpret Experiment 3a and pinpoint the mechanism behind the "curse of expertise." While participants in Experiment 3b acknowledged that their memory had decayed (the difference between Time 0 and Time 1), they were unaware of how much had actually been forgotten. Even after calling attention to the fact that "peak" knowledge is different than current knowledge, participants still did not accurately predict their ability to explain topics from their domain of formal expertise. People may be especially vulnerable to assuming that content that they had once mastered in great detail (e.g., knowing the full Krebs's cycle) degrades only slightly over time. By misestimating the rate of decline for their ability to produce explanations, participants rate themselves highly at Time 1 and only after having to confront their own ignorance through the writing task do they then lower their rating at Time 2.

7. General discussion

More knowledge can sometimes lead to greater metacognitive ignorance. People accurately self-assess in areas outside their expertise but are poorly calibrated within areas of expertise. Furthermore, education influences estimates of one's ability to explain. Although highly educated people do not exhibit the IOED for passive expertise, they do overestimate their ability to explain topics related to their formal expertise. This effect of education may be due to a misattribution of "peak" knowledge for current knowledge.

These studies reconcile two sets of findings. People with lower competence are the least aware of their level of competence (Dunning, 2005), and indeed those with less education do broadly exhibit the IOED for topics pertaining to passive expertise. Those who see themselves as experts show higher discrepancies between predicted and actual performance (Critcher & Dunning, 2009; Dunning et al., 2003; Glenberg & Epstein, 1987) and indeed those with more education do exhibit the IOED, but much more strongly in their domain of specialized study. Those with formal expertise exhibit meta-forgetfulness within their domain of knowledge, neglecting the rate at which deliberately learned information decays from memory.

Are the experts in our studies really oblivious to their explanatory ability, or are they providing high Time 1 ratings because they want to appear knowledgeable? Highly educated individuals are more likely to express strong opinions about non-existent public affairs topics (Bishop, Tuchfarber, & Oldendick, 1986), so perhaps participants think they can give high ratings and never be asked to verify the accuracy of their self-assessment. In the case of explanations this is most likely not the case. Even when participants were warned that they would be writing out a full explanation, they still showed the IOED (Rozenblit & Keil, 2002). Furthermore, if participants wanted to appear knowledgeable, they could have provided high Time 2 ratings as well. Instead, they lower their ratings indicating they are surprised by their lack of insight. We think our results are due to a genuine failure of meta-memory, which is only realized after actually attempting to produce a complete explanation.

Why is it that those with formal expertise accurately assessed their ability to explain domains of passive expertise? We did not directly explore this question in the current set of studies, but comparing the items used in our studies to the items used in related work may help provide an answer. Previous demonstrations of the IOED have used intermediately familiar items like a helicopter or a sewing machine (Rozenblit & Keil, 2002). Perhaps items like these are particularly difficult to accurately assess because they are neither familiar, which would lead to frequent recalibration, nor are they totally unfamiliar, which would lead to appropriate caution. Unlike previous demonstrations of the IOED, Experiment 1 sampled topics from the extremes that may have led to the improvement for those with formal expertise.

7.1. Future directions

The mechanism we propose is in the same vein, yet distinct from other memory-based heuristics posited to be contributors to overconfidence in fact-based knowledge. Cues such as current feeling states (Koriat, 1995), familiarity (Glenberg, Wilkinson, & Epstein, 1982; Metcalfe, Schwartz, & Joaquim, 1993), and fluency (Bjork, 1999) are assumed to be accurate indicators of memory quality (for review, see Metcalfe, 1998). The role of these cues in producing miscalibrated explanatory knowledge in domains of formal expertise warrants further research.

How can the illusion that one possesses “peak” knowledge persist? One possible explanation is that people fail to recognize the degree to which they are reliant on outside

sources to fill in the gaps of their knowledge. Merely searching the Internet, for example, can boost self-assessed knowledge (Fisher, Goddu, & Keil, 2015), so perhaps experts do not appreciate the division of cognitive labor upon which they rely and instead view themselves as islands of expertise with near-perfect memory for topics in their domain

Although in our experiments we used college education as a proxy for expertise, other forms of expertise like advanced degrees or years of experience in a field may lead to different effects. Perhaps these sorts of experts would have mastery of a domain to the point where there was minimal decay for explanatory knowledge, leading to well-calibrated self-assessments. The ability to explain could persist in cases where there is constant use and thus frequent updating and refreshing.

While our studies focused on the type of expertise as an individual difference variable, other associated individual differences may also be related to our findings. For example, reflective thought, as measured by the Cognitive Reflection Task (CRT), decreases the IOED (Fernbach, Sloman, Louis, & Shube, 2013) and, furthermore, CRT scores correlate with education (Welsh, Delfabbro, Burns, & Begg, 2011). Thus, cognitive ability construed as reflective thought might facilitate the pursuit of higher education and thus could be the more central variable of interest. It could also be closely linked to meta-forgetting. Alternatively, educational experience itself may increase levels of reflective thought. The specific influence of cognitive reflection on the curse of expertise remains an important topic for future work.

7.2. Conclusion

Explanatory reasoning is a pervasive aspect of cognition, yet people's intuitive theories are surprisingly incomplete. Here, we find that formal education does not consistently provide a remedy. While self-reported explanatory insight is higher in areas of expertise, the amount of miscalibration is also greater. This experimental evidence suggests that while education is associated generally with more accurate self-assessment, it also may create a blind spot in thinking about one's own knowledge: namely, one's area of concentrated deliberate study. Thus, expertise can influence self-assessment of explanatory ability, but in different ways as a function of the kind of expertise involved.

Acknowledgments

This work was supported by the John Templeton Foundation [IRES number: 13-003303]. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Fuller Thrive Center or the John Templeton Foundation.

Note

1. Low Education = without a 4-year bachelor's degree; High Education = 4-year bachelor's degree or more.

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.
- Arnold, L., Willoughby, T. L., & Calkins, E. V. (1985). Self-evaluation in undergraduate medical education: A longitudinal perspective. *Academic Medicine*, 60(1), 21–28.
- Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, 50(2), 240–250.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.
- Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society*, 17, 8–84.
- Camerer, C. F., & Johnson, E. J. (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections, and controversies* (pp. 342–364). Cambridge, UK: Cambridge University Press.
- Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97(6), 931–945.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (vol. 44, pp. 247–296). New York: Elsevier.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83–87.
- 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.
- Fernbach, P. M., Sloman, S. A., Louis, R. S., & Shube, J. N. (2013). Explanation friends and foes: How mechanistic detail determines understanding and preference. *Journal of Consumer Research*, 39(5), 1115–1131.
- Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, 144(3), 674–687.
- Fisher, M., & Keil, F. C. (2014). The illusion of argument justification. *Journal of Experimental Psychology: General*, 143(1), 425–433.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15(1), 84–93.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10(6), 597–602.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
- Herrmann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, 92(2), 434–452.
- Keil, F. C. (1998). Cognitive science and the origins of thought and knowledge. In R. M. Lerner (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (pp. 341–413). New York: Wiley.
- Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57, 227–254.

- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1–24.
- Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 686–715). Cambridge, UK: Cambridge University Press.
- Kominsky, J. F. & Keil, F. C. (2014). Overestimation of knowledge about word meanings: The “misplaced meaning” effect. *Cognitive Science*, 38, 1604–1633.
- Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124(3), 311–333.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 187–194.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280–286.
- Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, 1(1), 126–129.
- Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review*, 2(2), 100–110.
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 851–861.
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 502–506.
- Murphy, G. L. (2000). Explanatory concepts. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 361–392). Cambridge, MA: MIT Press.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289–316.
- Ozubko, J. D., & Fugelsang, J. (2011). Remembering makes evidence compelling: Retrieval from memory can give rise to the illusion of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 270–276.
- Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179.
- Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521–562.
- Sunderland, A., Watts, K., Baddeley, A. D., & Harris, J. E. (1986). Subjective memory assessment and test performance in elderly adults. *Journal of Gerontology*, 41(3), 376–384.
- Tracey, J. M., Arroll, B., Richmond, D. E., & Barham, P. M. (1997). The validity of general practitioners’ self assessment of knowledge: Cross sectional study. *BMJ. British Medical Journal*, 315(7120), 1426–1428.
- Welsh, M., Delfabbro, P., Burns, N., & Begg, S. (2011). Individual differences in anchoring: Numerical ability, education and experience. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3193–3198). Austin, TX: Cognitive Science Society.
- Wynne, G., Marteau, T. M., Johnston, M., Whiteley, C. A., & Evans, T. R. (1987). Inability of trained nurses to perform basic life support. *British Medical Journal*, 294, 1198–1199.
- Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, 66, 55–84.

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

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Training items

Appendix S2. Experiment 1 topics