The Illusion of Explanatory Depth in a Misunderstood Field: The IOED in Mental Disorders

Andrew S. Zeveney (asz214@lehigh.edu)
Department of Psychology, 17 Memorial Drive East
Bethlehem, PA 18015 USA

Jessecae K. Marsh (jessecae.marsh@lehigh.edu)
Department of Psychology, 17 Memorial Drive East
Bethlehem, PA 18015 USA

Abstract

Humans fail to understand the world around them and also fail to recognize this lack of understanding. The illusion of explanatory depth (IOED) exemplifies these failures: people believe they understand the world more deeply than they actually do and only realize that this belief is an illusion when they attempt to explain elements of the world. An unexplored factor of the IOED is how people may become overconfident by confusing their own understanding with others' understanding. In two experiments, we examine the IOED in mental disorders, a domain where society has a limited understanding. In Experiment 1, we demonstrate that people display an IOED for mental disorders as well as devices, but that it is smaller for mental disorders. In Experiment 2, we show that exposing the IOED is specifically linked to generating an explanation, rather than more generally thinking about a phenomenon.

Keywords: illusion of explanatory depth; domain differences; explanation

Introduction

Humans encounter a vast number of phenomena on a daily basis but only possess a shallow level of knowledge about most of these phenomena (Wilson & Keil, 1998). In addition to lacking an understanding of many everyday domains, people fail to recognize the limits of their own understanding, as demonstrated by the illusion of explanatory depth (IOED). First studied by Rozenblit and Keil (2002), the illusion of explanatory depth is the incorrectly held belief that one understands the world on a deeper level than one actually does.

The procedure used by Rozenblit and Keil to elucidate the IOED is simple. Participants were asked to rate their understanding of how devices worked (e.g., how a cylinder lock works). Next, for several of the items, they were asked to generate a detailed explanation of how each worked and to then rerate their understanding of that item. Consistently, ratings were lower after generating an explanation, suggesting that people do not understand the mechanisms of many everyday devices as much as they believe they do and only realize that they lack understanding after attempting to explain. The IOED has been shown in the domains of devices (Lawson, 2006; Mills & Keil, 2004; Rozenblit & Keil, 2002), natural phenomena (Rozenblit & Keil, 2002), and political policy (Alter, Oppenheimer, & Zemla, 2010; Fernbach, Rogers, Fox, & Sloman, 2013).

Why do people demonstrate an IOED? Explanations have typically centered on people's tendencies to overestimate the quality and depth of their mental representations and their failures to accurately judge their ability to provide good explanations (Mills & Keil, 2004; Rozenblit & Keil, 2002). Importantly though, people can and do rely on more than just their own understanding when generating explanations; they often seek out the understanding of others. People recognize that their own understanding has limits (even if they are inaccurate about where these limits lie) and use their understanding of what others know when choosing who to turn to fill in gaps in their own understanding (Keil, 2012; Wegner, 1987; Wilson & Keil, 1998).

Recent work suggests that because people so commonly rely on others for understanding, they may actually overestimate the amount of understanding that they possess when they cannot rely on others. For instance, Fisher, Goddu, and Keil (2015) demonstrated that people had greater confidence in their ability to explain various phenomena when they had first used the Internet to search for explanations for a different set of phenomena. They postulated that people's confidence in their own understanding came from conflating an understanding in the world with an understanding in their own mind. In another line of work by Kominsky and Keil (2014), people's illusory beliefs about their ability to generate differences in meaning between similar pairs of words were predicted by their beliefs that experts would be able to generate more differences. In general, this work suggests that the IOED may in part be confusing what others understand with what is understood in one's own mind.

Linking the IOED to the extent to which others understand a field has interesting implications for thinking about where an IOED might occur. For some domains, we as a collective society have a deep level of understanding of how things in that domain work. For example, there are people who understand devices like cylinder locks very well, even if the average layperson has a limited understanding of how a cylinder lock works. However, there are fields where understanding is fairly limited for everyone.

For instance, in the domain of mental health, there is limited understanding of how disorders should be defined (Barch & Keefe, 2009) or how treatments for mental health issues work (Wampold & Imel, 2015; Zhang & Malhotra, 2013). How would the IOED manifest in such domains where less is collectively understood about how things in this domain can be explained?

In the following set of experiments, we examine how our perceptions of what others understand about a domain influences the IOED for that domain. To explore this question, we examined the domain of mental disorders. Beyond the lack of scientific consensus on how symptoms interrelate in mental disorders to explain the etiology or presentation of a disorder (e.g., Matheson, Shepherd, & Carr, 2014), we have found that laypeople also recognize that understanding of mental disorders is limited. To measure this, we ran a pilot study where we asked a group of pretest participants (N = 50) to rate their beliefs about what we know about mental health, as well as devices, natural phenomena, and medical disorders. For each domain, participants rated their agreement that "we fully understand" the phenomenon (what we will call the society rating) on a scale of 1 (strongly disagree) to 7 (strongly agree). They also rated the difference between what an expert would know about each topic and what the average person knows about that topic (what we will call the gap rating) on a scale of 1 (no difference) to 7 (very great difference). We found that on average people reported higher societal understanding for devices (M = 5.73) than for mental disorders (M = 3.72; p < .001). People also reported a larger gap between lay and expert understanding for mental disorders (M = 5.64) than for devices (M = 3.73; p < .001). These findings provide evidence that laypeople recognize the difficulty we have in understanding mental disorders on a deep level. With these findings in mind, mental disorders serve as a prime domain in which to examine several questions about the illusion of explanatory depth.

Research Questions

Question 1 Do people demonstrate an IOED even in a field where understanding is limited?

Since people recognize that others, particularly laypeople, understand fairly little about mental health, one might anticipate that people will be more accurate about their own understanding. However, the IOED is a robust effect found in multiple domains, so people may still hold an illusion even when they recognize that there is limited understanding overall. The size of this illusion may be smaller in mental health than devices because people endorse a lower level of understanding for mental health at the outset.

Question 2 Do people's individual beliefs about others' understanding predict their own illusion?

If an individual's own beliefs about what others understand about a domain predicts the magnitude of his or her IOED for that domain, it would lend support to the hypothesis that the IOED occurs in part because people believe that understanding exists in the world and mistakenly believe that they have a greater portion of this understanding than they actually do.

Question 3 Is explanation necessary to expose the illusion?

Finally, mental health serves as a domain in which to further test the theory that it is specifically explanation that exposes the illusion (Rozenblit & Keil, 2002). Examining this is particularly important in mental health, as there is plenty of information about things like risk factors and symptoms but limited ability to explain how the interplay of these aspects leads to disorder. Is an attempt to explain necessary to reveal the illusion or will thinking about one's knowledge of disorder be sufficient?

Overview of Experiments

In Experiment 1, we presented participants with a number of phenomena including mental disorders and devices and asked them to rate their understanding before and after attempting to explain each phenomenon. Additionally, we asked participants to rate their beliefs about what others understand. In Experiment 2, we asked participants to rate their understanding of mental disorders but varied whether they were asked to explain the disorder or describe features of each disorder.

Experiment 1

In Experiment 1, we examined the IOED in mental health items using the basic protocol established by Rozenblit and Keil (2002). If people do have an illusion of their own understanding for mental disorders, we would expect to see a drop in ratings after attempting to explain these items. We also tested the IOED in devices in order to have a baseline to which to compare the potential magnitude of the illusion in mental health.

We also examined the correlations between beliefs about the general level of understanding and ratings of personal understanding. If people confuse their own understanding with what others understand, we anticipated that perceptions of the amount of societal understanding would be correlated with higher perceived personal understanding. Additionally, we examined the correlations between beliefs about the gap between lay and expert knowledge and ratings of personal understanding. We anticipated that a perception of a smaller gap between lay and expert knowledge would allow for more confusion between personal understanding and others' understanding, given that understanding is more widely spread across individuals, rather than being concentrated in experts. Thus, we expected that perceptions of a smaller gap would be correlated with higher ratings of personal understanding.

Methods

Participants 150 participants recruited through Amazon's Mechanical Turk participated for payment.

Materials and Procedure Participants were assigned to one of three conditions: devices (n = 50), mental disorders (n = 49), or mental health treatments (n = 51). The mental health treatment condition was included for comparisons not presented here and will not be discussed further. Participants were presented with a five-item stimuli set consisting of items from the given domain (see Table 1 for examples of all stimuli).

Table 1
Example stimuli in each domain

Domain	Stimuli
Devices	How a zipper works
Mental disorders	How the different symptoms of depression develop
Mental treatments	How an antipsychotic works

Participants were instructed on how to rate their understanding using a 1 (low) to 7 (high) understanding scale adapted from Rozenblit and Keil (2002). They were given an example of a high, moderate, and low level of understanding. They then rated the stimuli set on this scale (Time 1 [T1] rating). Following, participants were again presented with the five items, one by one, and prompted to explain each item in as much detail as possible. After each explanation, participants rated their understanding of that phenomenon on the understanding scale (Time 2 [T2] rating).

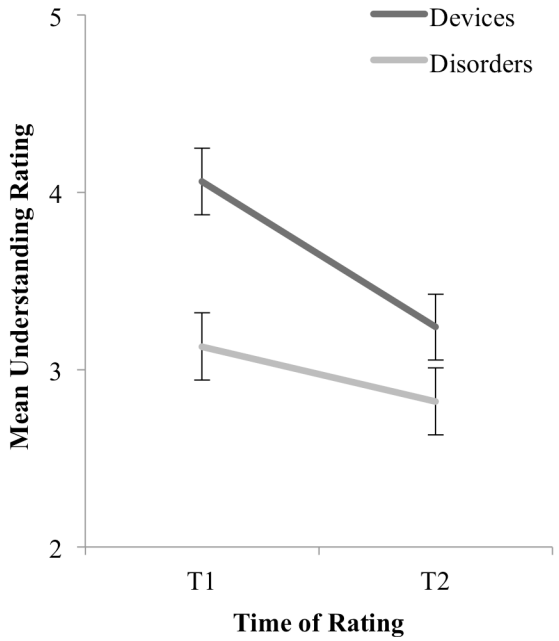
Participants then answered several follow-up questions designed to assess beliefs about others' understanding of each domain. For each of the items in the original set, participants rated their agreement that "we fully understand" the phenomenon ("society rating", rated from 1, strongly disagree, to 7, strongly agree). They also rated the size of the difference between what an expert would know about each topic and what the average person knows about that topic ("gap rating," rated from 1, no difference to 7, very great difference). Additionally, participants answered basic demographic questions.

Results

We examined ratings for devices and mental disorders using a 2 (time: T1 vs. T2; within) x 2 (domain: devices vs. mental disorders; between) mixed ANOVA. We found a significant effect of time, $F(1, 97) = 40.96$, $p < .001$, $\eta_p^2 = .30$, with ratings of understanding decreasing from T1 (M = 3.59, SD = 1.40) to T2 (M = 3.03, SD = 1.33). We also found a significant effect of domain, $F(2, 97) = 7.23$, $p = .008$, $\eta_p^2 = .07$. Overall, participants rated their understanding of devices (M = 3.65, SD = 1.24) as significantly higher than their understanding of disorders (M = 2.98, SD = 1.25).

Finally, there was a significant interaction between time and domain, $F(2, 97) = 3.41$, $p = .036$, $\eta_p^2 = .044$. Sidak-corrected follow-up tests showed the expected illusion in devices, as ratings significantly dropped from T1 to T2, $p < .001$. Additionally, people also demonstrated an illusion in mental disorders, $p = .014$, as ratings dropped significantly from T1 to T2 in this domain as well (see Figure 1). In order to compare the size of the illusion across domains, we examined this interaction another way. At T1, ratings for devices were significantly higher than those for mental disorders, $p = .001$. However, at T2, ratings did not differ between devices and mental disorders, $p = .112$. Given that ratings for devices were greater than mental disorders at T1 but the same at T2, participants showed a greater drop (and thus a larger illusion) for devices than for mental health phenomena.

Figure 1: The IOED in devices and mental disorders



Next, we examined the society and gap ratings. Participants rated society as understanding devices, M = 4.42, SD = 1.57, significantly more than mental disorders, M = 3.55, SD = 1.41, $t(97) = 2.92$, $p = .004$. Participants also rated the gap between expert and lay knowledge as significantly larger for disorders, M = 5.71, SD = 1.15, than for devices, M = 4.84, SD = 1.10, $t(97) = 3.88$, $p < .001$. These findings replicated our pilot data.

We then examined the correlations between society/gap ratings and ratings of personal understanding at T1 and T2 across domains. Society ratings were correlated with personal understanding of disorders at T1 ($r = .310$, $p = .032$) and T2 ($r = .400$, $p = .005$). For devices, society ratings were also correlated with personal understanding at

T1 (r = .290, p = .043) and T2 (r = .412, p = .003). Perceptions of higher societal understanding were related to higher individual understanding both before and after generating an explanation.

For gap ratings, a slightly different pattern emerged. For mental disorders, smaller perceived gaps were correlated with higher personal understanding for mental disorders at T1 (r = -.315, p = .029) and T2 (r = -.491, p < .001). However, gap ratings were not correlated with personal understanding ratings for devices (rs < .036, ps > .810). Thus, the anticipated relationship between smaller gaps and higher personal understanding was found only in mental disorders.

Discussion

As in devices, people demonstrate an illusion of explanatory depth in the domain of mental disorders, even though they endorse less general understanding and in particular, less lay understanding of disorders than devices. This finding serves to validate the robustness of the IOED: even when people admit that a field is not particularly well understood and that understanding is more concentrated within experts, they still overestimate their own understanding.

Importantly, the illusion was larger in devices than in mental disorders, as personal understanding was higher at T1 for devices than for mental disorders but not at T2. Potentially, one explanation is that people confuse their own understanding of a phenomenon with the understanding that others have for that phenomenon. Given that people see mental health as less understood by others, people may initially experience less confusion between others' understanding and their own in mental disorders than in devices. As a result, ratings of personal understanding for mental disorders at T1 are closer to actual levels of understanding, leading to a smaller illusion in mental disorders than devices.

Relatedly, the illusion in mental disorders may also be smaller because people see understanding as more concentrated within experts for disorders than for devices. People may be less likely to confuse their understanding with others' understanding when others' understanding is seen as mostly possessed by experts on a domain, rather than being distributed more evenly among all people. Future research should explore perceptions of both the level of understanding that others possess and the distribution of this understanding across others.

Additionally, this experiment sheds light on the potential relationships between individuals' perceptions of others' understanding and of their own understanding. For mental disorders, perceptions of societal understanding and the lay/expert gap were related to personal understanding at both T1 and T2, and for devices, societal and personal understanding showed the same relationships. Importantly, the same relationships were found both before generating an explanation and after, when ratings of personal understanding are more accurate. Thus, it is not that individuals' beliefs about what society understands and about the differences between expert and lay understanding simply relate to their illusion of understanding: these beliefs are correlated to perceptions of understanding even after the illusion is shattered through explanation. One thing that is important to note is that society and gap ratings were completed only after participants had generated an explanation. Potentially, generating this explanation altered beliefs about what others understand and therefore, obscured the true relationship between beliefs about others' understanding and personal understanding.

Given that the IOED in mental disorders is smaller than in devices, another possible question arises of whether this illusion is similar in nature to the IOED shown in other domains. Prior work on the IOED has suggested that it is only through attempting to explain a phenomenon that one is made aware of one's own limits in understanding (Rozenblit & Keil, 2002). But when considering many phenomena, including mental disorders, people can easily think about these phenomena in ways that do not entail detailed explanations (e.g., thinking about all of the symptoms of a disorder without thinking about how they are interrelated to create an explanation of how the disease came into existence). Is this type of deeper thinking sufficient to expose the illusion or is it actually necessary to generate an explanation? We tested this possibility to determine if the IOED in mental health disorders is reliant on explanations as it is in other domains.

Experiment 2

In Experiment 2, we tested whether explanation is specifically necessary to expose the IOED, or whether simply thinking deeply about a phenomenon more generally is enough. We also tested a wider range of mental disorders, in order to further extend the finding that people do demonstrate an IOED for mental health issues.

Methods

Participants 100 participants recruited through Amazon's Mechanical Turk participated for payment.

Materials and Procedure

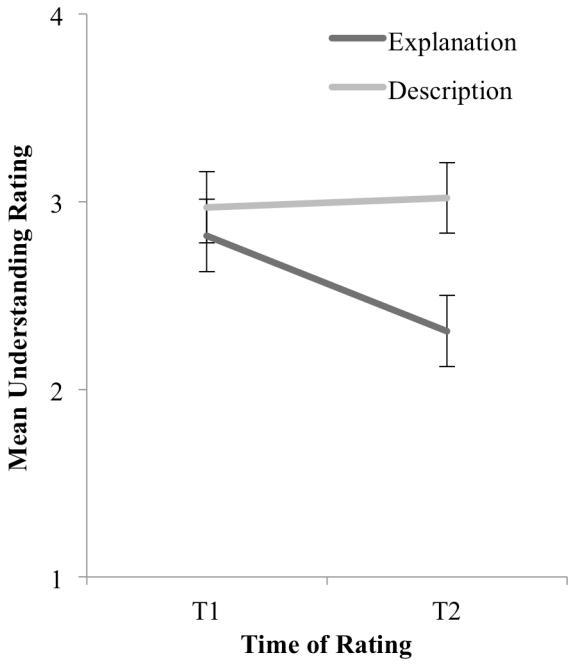
Participants were assigned to either the explanation condition (n = 50) or the description condition (n = 50). All participants were presented with eight mental disorders: depression, OCD, generalized anxiety disorder, schizophrenia, borderline personality disorder, bipolar disorder, ADHD, and anorexia. For a given disorder, participants were presented with the name of a disorder and then were asked to rate their understanding with the same question as in Experiment 1 (T1 rating). Participants in the explanation condition then were asked to explain each disorder as in Experiment 1. Participants in the description condition were asked to list as many characteristics about

the disorder as they could. After completing the explanation or the description task, participants completed the T2 rating from Experiment 1. Participants then completed the society and gap ratings from Experiment 1. Finally, participants rated each disorder in terms of how biologically, psychologically and environmentally caused they were on a scale of 0 (not at all) to 100 (completely). Causation questions were not the focus of the work presented here, so they will not be discussed further.

Results

We examined ratings using a 2 (time: T1 vs. T2) x 2 (condition: explanation vs. description) mixed ANOVA. We found a significant effect of time, $F(1, 98) = 6.79$, $p = .011$, $= .065$, and a marginally significant effect of condition, $F(1, 98) = 2.82$, $p = .011$, $\eta_p^2 = .065$. These main effects should be interpreted in light of a significant interaction between time and condition, $F(1, 98) = 4.12$, $p = .002$, $\eta_p^2 = .097$. Sidak-corrected follow-up tests showed that in the explanation condition, there was a significant drop from T1 ($M = 2.98$, $SD = 1.33$) to T2 ($M = 2.31$, $SD = 1.27$; $p < .001$). However, for the description condition, there was not a significant difference between T1 ($M = 2.97$, $SD = 1.33$) to T2 ($M = 3.02$, $SD = 1.30$; $p = .650$; see Figure 2).

Figure 2: Experiment 2 mean understanding ratings.



Examining this interaction another way, ratings between conditions did not differ at T1, $p = .454$, but are significantly lower at T2 for the explanation condition than the description condition, $p = .002$. Thus, participants in both conditions endorsed the same initial level of understanding but only those who explained their understanding showed a drop at T2. The illusion was only revealed when people attempted to explain mental disorders, and not when they simply described their knowledge.

We also examined the correlations between society/gap ratings and personal understanding. For the description condition, society ratings were correlated with T1, $r = .370$, $p = .008$, and T2 ratings, $r = .408$, $p = .003$, and gap ratings were uncorrelated with T1 ratings, $r = -.171$, $p = .234$, and marginally correlated with T2 ratings, $r = -.269$, $p = .059$. In the explanation condition, participants showed a different pattern. Society ratings were uncorrelated with T1 or T2 ratings, rs > .190, ps > .135, and gap ratings were correlated with T1, $r = -.294$, $p = .042$, and T2 ratings, $r = -.316$, $p = .029$. Higher ratings of societal understanding were related to higher ratings of personal understanding but only in the description condition. Likewise, ratings of a larger lay/expert gap were negatively related to personal understanding only in the explanation condition.

Discussion

As in Experiment 1, participants demonstrated an IOED for mental disorders. However, this illusion was only exposed when people attempted to explain disorders, not when they were simply prompted to think about the aspects of the disorder. This finding provides further evidence that overconfidence in one's understanding can be revealed by attempting to explain that understanding, but by simply reflecting on aspects of the phenomenon that might be related to that understanding.

In this experiment, we also found that the relationships between perceptions of others' understanding and personal understanding differed across conditions. Personal understanding was correlated with gap ratings only in the explanation condition and with social understanding only in the description condition. This finding provides further evidence that generating an explanation changes people's later beliefs about others' understanding. The conditions were only different from one another after T1 ratings were made, so relationships for T1 ratings and others' understanding should be the same if perceptions of others' understanding are not changed by the experiment. Despite the difficulty of interpreting these results, overall our results suggest that ratings of societal understanding and the lay/expert gap are related to individuals' ratings of understanding.

General Discussion

Across two studies we explored whether an illusion of explanatory depth is demonstrated for a domain where people acknowledge that we have limited understanding of the domain. Our pretest study indicated that laypeople do in fact believe that we as a society have less of an understanding of mental disorders than devices and understanding is more concentrated within experts in the

mental health domain. In comparing the IOED across mental health disorders and devices, we saw a smaller IOED in disorders than devices (Experiment 1). This IOED seemed to be based in the same mechanism as previous research (Rozenblit & Keil, 2002) in that it only arose after explanation and not after mere description (Experiment 2). Despite the seeming link between mental disorders being seen as less well understood and a smaller IOED in the mental domain, ratings of others' understanding were correlated not only with initial illusory perception of personal understanding but also with personal understanding after generating an explanation.

Finding that people show a smaller IOED for mental health than devices suggests that though people are overconfident about their understanding in a number of domains, the extent to which they are overconfident may vary. In particular, mental health may be a domain in which people are more accurate about their understanding, and this accuracy may be driven by their perception that this domain is less understood in general. At the same time, there is still a gap between what people believe that they understand before and after attempting to explain.

Our findings suggest avenues for future research related to the IOED. For example, understanding what particular domain differences lead to the difference in size of the IOED across domains would be an important avenue to explore to better understand the underlying mechanism of the IOED. There may be certain domains in which laypeople would be unwilling to endorse any understanding, believing that only an expert can understand that particular subject. How unfamiliar would a domain need to be to laypeople for them to accurately assess their limited understanding? These are questions for future research that deserve exploring.

Furthermore, our research has interesting implications for how people reason about the mental health domain. It is interesting to think about how generating explanations as participants did in this study may alter how people think about their own diagnoses or treatments, and those of the people around them. For example, even if people are aware that they do not completely understand how the disorder of depression comes to be, our results suggest that they would still overestimate how much they understand about the disorder in general. This overconfidence may give people a false sense of understanding what treatment would best address a disorder's symptoms since they feel like they understand how the disorder works at least in some part. It is an interesting avenue for future research to explore how these false senses of understanding influence decision making in health.

Overall, our results are informative as to the extent of people's overconfidence in their own understanding. Even in the domain of mental disorders, where causal links are complex, multifaceted, and invisible, people still endorse a greater understanding than they actually possess. While it seems that in some ways, this illusion is supported by a belief that others in the world understand the domain, the relationship between our own understanding and others' understanding is potentially more complicated and deserves future research.

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*, 436-451.

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science, 24*, 939-946.

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*, 674-687.

Kominsky, J. F., & Keil, F. C. (2014). Overestimation of knowledge about word meanings: The "misplaced meaning" effect. *Cognitive Science*, *38*, 1604-1633.

Lawson, R. (2006). The science of cycology: Failures to understand how everyday objects work. *Memory & Cognition*, *34*, 1667-1675.

Matheson, S. L., Shepherd, A. M., & Carr, V. J. (2014). How much do we know about schizophrenia and how well do we know it? Evidence from the Schizophrenia Library. *Psychological Medicine*, *44*, 3387-3405.

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

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

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: Research evidence for what works in psychotherapy* (2nd ed*.)*. New York: Routledge.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York, NY: Springer-Verlag.

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

Zhang, J. P., & Malhotra, A. K. (2013). Pharmacogenetics of antipsychotics: Recent progress and methodological issues. *Expert Opinion on Drug Metabolism & Toxicology*, *9*, 183-191.