



The Influence of Exposing the Illusion of Explanatory Depth on Communication

Allison Seohyun Hong

Abstract

The Illusion of Explanatory Depth (IOED) is a type of cognitive bias that causes humans to overestimate their understanding of mechanisms. When individuals believe they understand a subject well, their confidence spikes without realizing their true shallow knowledge. Research proves that the IOED is why extreme polarization in politics happens. Proving a point thrives on the confidence of the individual; therefore, it is crucial to ask the right questions in the discussion. Humans, being cognitive misers, tend to look for shortcuts to store information in their brain, surviving around a concept called the community of knowledge. Arguments are usually combat-based, where each side of the party will do anything to get their point across. However, exposing the Illusion of Explanatory Depth can cause arguments to shift to a more cooperative discussion. When individuals are asked to explain the underlying mechanics of whatever they are supporting, it leaves them in conflict where they realize they have been acting on their own arrogance, rather than true information. The exposition ultimately raises intellectual humility and causes both parties to lower their defenses, leading to a productive discussion. While this exposition does not act universally, it is significant to humble disagreements with the IOED to improve personal communication.

1. Introduction

People rarely hesitate to defend their opinions. Confidence fuels them and almost always the loudest, most arrogant person wins the conversation. Yet when asked to explain exactly how the systems or the mechanics behind those opinions work, their own confidence often gives way to uncertainty. This phenomenon is called the Illusion of Explanatory Depth (IOED) where humans mistake their surface level familiarity for a deep understanding (Rozenblit & Keil, 2002). This cognitive bias was first discovered during a study in 2002 by Leonid Rozenblit and Frank Keil. They asked participants to rate their understanding of everyday objects, such as zippers, before asking them to elaborately explain the simple mechanisms of the objects. They were then asked to rerate their own understanding, which was substantially lower after. Rozenblit and Keil's conclusion was that asking humans to explain simple concepts confronts them with the truth that they do not understand as much as they think they do. However, this does not just occur with understanding of objects but also with opinions and policies. In modern conversations and debates, people with the desire for power passionately defend complex social, political, or philosophical viewpoints without actually understanding what they are arguing about (Körner et al., 2024).



In 2013, researcher Philip Fernbach and other scientists conducted a study where they discovered that this cognitive blind spot is fueling modern political extremism (Fernbach et al., 2013). Fernbach found that individuals with more extreme political positions often overestimated how well they understood these policies. When they were required to explain how these systems worked, it reduced the extremity of their views. People scream about complicated laws or social programs based on pure emotion thinking they know this subject from the bottom of their hearts without realizing their brains are running on a mirage (Voelkel et al., 2018). Yet, researchers identified a strategy that could reduce this overconfidence: when these extreme participants were forced to stop listing why they were so passionate about their position and instead explain the exact process of how their policy would be successful, their willingness to be moderate rose. By confronting their own ignorance, their expression became less extreme (Chesterfield & Coombs, 2019). This paper investigates how exposing the Illusion of Explanatory Depth influences personal communication, arguing that gently forcing individuals to confront their own knowledge is a powerful tool for lowering defensive behavior and shifting arguments from a combat mindset to a more cooperative one.

2. Cognitive Misers

Human brains are the perfect foundation for this kind of phenomenon to happen. We are cognitive misers, where when given two choices between tasks, we almost always immediately decide on the option that takes less effort and time (Nador et al., 2020). The human brain likes to save processing energy, always looking for a shortcut for objectives. This includes memory as well. When we look back on a familiar topic, the brain views it from a high abstract view, making it so that it is easier to grasp and we get the feeling that we understand it. Thinking deeply is a lot of work for the brain and it takes a lot of energy and effort to do so. Moreover, it is easier to avoid the intense mental work required to store complex blueprints of how things work. Therefore, humans survive by participating in the “community of knowledge”, where knowledge is distributed socially. Humanity works as a collective group that understands how a smartphone is built or how some policy is managed, which results in individual human brains outsourcing that data. Humans do not need to memorize how a toilet functions because all we do is turn the flush handle. So, IOED occurs because the brain is terrible at tracking where its personal knowledge ends and the community’s knowledge begins. We confuse our ability to access information with actually possessing it. Crucially, people generate reasons much more easily than mechanistic explanations (Alter et al., 2010). The human brain is incredibly fast at generating emotional reasons for why it likes an idea or a side in a debate. However, the brain struggles with tracing the actual mechanisms of how that idea functions (Mills & Keil, 2003). Therefore, when an individual is forced to provide a step-by-step breakdown, the information that is outsourced from the community is missing, which may produce an immediate phenomenon called cognitive dissonance: the uncomfortable mental state where your actions

conflict with your beliefs (*Cognitive Dissonance* | *Psychology* | *Research Starters* | *EBSCO Research*, n.d.). Our confident ego is confronted by our shallow understanding, where we are not exactly sure how something works but proceed to arrogantly talk about it anyway. This dissonance causes us to feel unsure about what we are saying all of the sudden and when thought about in the middle of an argument, it can significantly reduce confidence and put us in an awkward position.

3. Consequences of Exposition

When the Illusion of Explanatory Depth is exposed within personal conflicts, it triggers a shift in human interaction. The dismantling of this illusion serves as a dynamic foundation that transforms social behavior. By forcing someone to confront the real depth of their knowledge, the individual can alter the entire psychology of an interaction (Mehta & Mehta, 2024). This change falls across distinct behavioral phenomena: the elevation of intellectual humility and the structural modification of argumentative situations.

The first psychological effect of exposing the IOED is an immediate rise in intellectual humility, which is the awareness of one's limitations in their own knowledge. Under normal conversations, humans present themselves with as much confidence and mastery as they can to try and get the upper hand. This unearned confidence essentially acts as an armor, causing these individuals to become highly defensive when someone opposes their viewpoint. However, when they are forced to generate a strict elaboration, they experience a rapid decrease in cognitive confidence, which alters how the human ego evaluates its own capability.

Research suggests that exposing the Illusion of Explanatory Depth may have broader effects beyond a single topic (Meyers et al., 2023). Researchers from Cambridge University found that after participants attempted to explain the mechanics of one phenomenon, they subsequently rated their understanding of other unrelated topics less as well. While one might assume that realizing you do not completely understand the inner mechanism of a specific system would only make you humble about that one specific topic, research says otherwise and proves that this is highly systemic. Just like how emotions bleed over to other domains of one's life, when people are exposed in one area of expertise, their brain undergoes a reset which results in feeling outed in other fields. Therefore, in the initial experiment of the IOED, the participants lowered their self report knowledge scores across all completely unrelated topics to the original object they were asked about. This spillover effect can alter just how a conversation can flow now that the intellectual humility of a person is increased. Overconfident people look for information that confirms what they think they know. When they are first exposed to their knowledge and forced into a reset, the reduced confidence may increase one's openness to new information. Since they are not as defensive anymore, their brains stop blocking opposing evidence and instead

begin to search for new data to fill their newly discovered mental gaps. Not only does this make communicating a lot easier, but it helps people to learn more and expand their perspective. Similarly, once the spillover effect introduces intellectual humility, they become less concerned with defending their image and instead more willing to acknowledge these gaps (Porter et al., 2022). Because the person has already accepted that they have boundaries to their knowledge, they no longer view the disagreement as a threat but as an opportunity to build a more accurate foundation. Research on intellectual humility also proves that recognizing one's limitations is associated with greater openness to learning while also dismantling defensiveness. Moreover, studies on different types of question asking indicate that people often overestimate how incompetent they will appear when admitting they do not know something (Cojuharenco & Karelaia, 2019).

In everyday human interaction, arguments are backed by a combat mindset. In this state, individuals view the discourse as a win-or-nothing game where the main objective is to assert dominance on their viewpoint rather than seeking mutual understanding. They see all other opposing viewpoints as a threat and get highly defensive. When individuals are asked why they hold an opinion, they can easily produce highly emotional and surface level moral justifications. And when they do that, it reinforces their overconfidence and shuts down their active listening. To avoid this, it is important to shift the questions from following "why" to "how." In the experiments done by Fernbach and his team, participants were randomly assigned to one of two conditions. In the reason-generation condition, they listed the reasons why they supported or opposed a public policy. In the explanation condition, they were asked to explain how a policy would actually work when in practice. There was a contrast between these two groups: Participants who generated reasons maintained their perceived understanding and reinforced the strength of their original position. Those asked to explain the policy's mechanics, however, quickly encountered gaps in their knowledge. As these participants struggled to construct a reasonable explanation, their self rated understanding declined and they began talking in a tone that was more moderate than passionate. Howard Becker of University of Chicago also studied this concept and found out that asking people why they supported or did something made them more defensive (Allspaw, 2020). The interviewees asked "why" mistook the interviewer's question as a demand for a justification instead of seeing it as a pure question of curiosity. So if asked for justification, the response would be crafted so it would look good on their part. However, if asked "how", the interviewee had to create a response based on logical thinking. Not only does this method set a foundation for a more cooperative conversation, it causes the person to reflect on their reasoning and to find out whether they truly know a subject or not. By manipulating a conversation so a person switches from abstract reasons to logical explanations, the entire nature of the disagreement changes. The other person is disarmed of its defensive status by the sudden confrontation that they may have a gap in what they know. The interaction

then shifts from a combat zone where everyone is overconfident to a cooperative discussion where both parties look at the actual facts of the problem.

4. Resistance to Exposition

While exposing the IOED is a powerful tool for deflating overconfidence and promoting productive conversations, evidence suggests it does not work as a universal cure. The effectiveness of this confrontation is held back by several limitations, most significantly when one's beliefs are tied to an individual's core identity or when certain traits dominate. When a belief is not a preference but a structure of an individual's moral or religious identity, confronting them with gaps in their knowledge does not produce intellectual humility, but rather threatens them even more (Kahan, 2017). This concept is called identity-protective cognition, where people would go as far as forcibly believing misinformation that affirms their identity rather than threatening it. Although early research proposed that such challenges could produce a backfire effect, in which these corrections strengthen the false beliefs even more, recent studies suggest that this response is uncommon (Wood & Porter, 2018). Therefore, bringing up objective facts to try and prove the other wrong will be ineffective (Gollini, 2021). Instead, resistance to changing one's views is consistently explained by identity-protective cognition and motivated reasoning, both of which make individuals more likely to defend beliefs that are central to who they are.

Furthermore, individual personality differences also dictate how a person reacts when they are exposed to the IOED. People with high narcissism, low empathy and a low tolerance for cognitive dissonance decrease the likeliness of the spillover effect happening. For these people, admitting that they don't actually know what they're talking about is viewed as an entire defeat. It does not shift the conversation from combative to cooperative, but may end up resulting in the person employing their defense mechanisms to preserve their fragile ego. Therefore, while exposing the IOED is highly effective for moderate or open minded people, it's crucial to be extremely cautious when engaging with highly defensive personalities and individuals who deeply stand their ground.

5. Conclusion

In conclusion, the human mind is a master of illusion, constantly mistaking the convenience of holding knowledge for the ability of generating it. We navigate a highly complex world by borrowing the collective understanding of the community around us, but our brains lazily take the credit for it. While this cognitive miser shortcut is highly efficient, it becomes a dangerous tool where when used to an extent, it can raise arrogance, polarization, and defense in our daily social discourse.

This paper has investigated the transformative influence of exposing the Illusion of Explanatory Depth within personal communication. By shifting the prompt of a disagreement from “why” to “how”, we can gently guide individuals to the real boundaries and limits of their own understanding. This results in the raised intellectual humility in others, which helps conversations shift from combat-focused to cooperative, becoming a foundation for more productive dialogue. Ultimately, true communicative influence is about finding courage to accept our brain’s ignorance, rather than projecting unearned overconfidence to undermine others.

References:

- [1] Allspaw, J. (2020, January 8). “*Tricks of the Trade*” on ‘how?’ versus ‘why?’ <https://www.adaptivecapacitylabs.com/2020/01/08/tricks-of-the-trade-on-how-versus-why/>
- [2] 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>
- [3] Chesterfield, A., & Coombs, K. (2019, November 25). To fight polarization, ask, “How does that policy work?” *Behavioral Scientist*. <https://behavioralscientist.org/to-fight-polarization-ask-how-does-that-policy-work/>
- [4] *Cognitive dissonance | Psychology | Research Starters | EBSCO Research*. (n.d.). EBSCO. <https://www.ebsco.com/research-starters/psychology/cognitive-dissonance>
- [5] Cojuharenco, I., & Karelaia, N. (2019a). When leaders ask questions: Can humility premiums buffer the effects of competence penalties? *Organizational Behavior and Human Decision Processes*, 156, 113–134. <https://doi.org/10.1016/j.obhdp.2019.12.001>
- [6] Deng, Z., Deng, Z., Deng, Z., & Deng, Z. (2025). Becoming a cognitive miser? Antecedents and consequences of addictive ChatGPT use. *Social Science & Medicine*, 383, 118467. <https://doi.org/10.1016/j.socscimed.2025.118467>
- [7] 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>
- [8] Gollini, S. (2021, May 31). Backfire effect: why debunking does not regenerate our thinking. *Nexus*. <https://nexusconsultation.com/en/2021/05/31/backfire-effect-why-debunking-does-not-regenerate-our-thinking/>
- [9] Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of Identity-Protective Cognition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2973067>
- [10] Körner, R., Schütz, A., & Petersen, L. (2024). “It doesn’t matter if you are in charge of the trees, you always miss the trees for the forest”: Power and the illusion of explanatory depth. *PLoS ONE*, 19(4), e0297850. <https://doi.org/10.1371/journal.pone.0297850>

- [11] Mehta, S., & Mehta, N. (2024). Embracing the illusion of explanatory depth: A strategic framework for using iterative prompting for integrating large language models in healthcare education. *Medical Teacher*, 47(2), 208–211. <https://doi.org/10.1080/0142159x.2024.2382863>
- [12] 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. <https://doi.org/10.1017/jdm.2023.24>
- [13] Mills, C. M., & Keil, F. C. (2003). 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), 1–32. <https://doi.org/10.1016/j.jecp.2003.09.003>
- [14] Nador, J., Harel, A., Juvina, I., & Minnery, B. (2020). The Case of the Cognitive F(Opti)miser: Electrophysiological Correlates of Working Memory Maintenance Predict Demand Avoidance. *Journal of Cognitive Neuroscience*, 32(8), 1550–1561. https://doi.org/10.1162/jocn_a_01566
- [15] Porter, T., Elnakouri, A., Meyers, E. A., Shibayama, T., Jayawickreme, E., & Grossmann, I. (2022). Predictors and consequences of intellectual humility. *Nature Reviews Psychology*, 1(9), 524–536. <https://doi.org/10.1038/s44159-022-00081-9>
- [16] Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1
- [17] Voelkel, J. G., Brandt, M. J., & Colombo, M. (2018). I know that I know nothing: Can puncturing the illusion of explanatory depth overcome the relationship between attitudinal dissimilarity and prejudice? *Comprehensive Results in Social Psychology*, 3(1), 56–78. <https://doi.org/10.1080/23743603.2018.1464881>
- [18] Wood, T., & Porter, E. (2018). The elusive backfire effect: mass attitudes' steadfast factual adherence. *Political Behavior*, 41(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>