



Do It Yourself Content and the Wisdom of the Crowds

Dallas Amico-Korby¹ · Maralee Harrell² · David Danks³

Received: 10 July 2024 / Accepted: 20 December 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract

Many social media platforms enable (nearly) anyone to post (nearly) anything. One clear downside of this permissiveness is that many people appear bad at determining who to trust online. Hacks, quacks, climate change deniers, vaccine skeptics, and election deniers have all gained massive followings in these free markets of ideas, and many of their followers seem to genuinely trust them. At the same time, there are many cases in which people seem to reliably determine who to trust online. Consider, for example, Do It Yourself (DIY) content about how to play guitar, bake, fix one's plumbing, or repair one's car. For these topics, those who have the largest accounts and the most popular content typically possess significant expertise. That is, social media users seem to reliably pick out DIY experts. We thus have a puzzle: why are social media users seemingly competent at identifying DIY experts, but not climate science or vaccine experts? In what follows, we solve this puzzle. We begin by identifying a novel wisdom of the crowds phenomenon: specifically we argue that the crowd (in combination with social media search and recommendation algorithms) reliably picks out DIY experts and serves as a credentialing institution for DIY content. Next, we argue that (a) there are five epistemic factors that determine whether the crowd can succeed at recognizing experts on social media platforms, and (b) while many of those factors are satisfied to a sufficiently high degree by DIY content, they are mostly lacking for content about climate change or vaccines.

✉ Dallas Amico-Korby
damico@ucsd.edu

Maralee Harrell
mharrell@ucsd.edu

David Danks
ddanks@ucsd.edu

¹ UC San Diego Philosophy, San Diego, CA, USA

² UC San Diego Philosophy and School of Public Health, San Diego, CA, USA

³ UC San Diego Data Science and Philosophy, San Diego, CA, USA

1 Introduction

Many social media platforms enable (nearly) anyone to post (nearly) anything. One clear downside of this permissiveness is that many people appear bad at determining who to trust online. Hacks, quacks, climate change deniers, vaccine skeptics, and election deniers have all gained massive followings in these free markets of ideas, and many of their followers seem to genuinely trust them (as evidenced by vaccine misinformation being negatively correlated with intent to vaccinate (Enders et al., 2020; Loomba et al., 2021; Latkin et al., 2023), or by followers showing up at the White House on January 6th (Van Swol et al., 2022)).

At the same time, there are many cases in which people seem to reliably determine who to trust online. Consider, for example, Do It Yourself (DIY) content about how to play guitar, bake, fix one's plumbing, apply cosmetics, or repair one's car. For these topics, those who have the largest accounts and the most popular videos also possess significant expertise. That is, social media users seem to reliably pick out some (though not necessarily all) experts in these areas.

We thus have a puzzle: why are social media users seemingly competent at identifying DIY experts, but not climate science or vaccine experts? DIY content (e.g., guitar tutorial content) and content about vaccines (for example) are similar in many respects. Platforms typically treat them similarly, allowing nearly anyone to post nearly anything without platform-generated assessments of expertise or accuracy.¹ In fact, to the extent that there is a difference in how they are treated, content about vaccines is typically held to *higher* standards since it is often more important that people have accurate information. Moreover, all these topics, from guitar playing to virology, have posts and content from both experts and non-experts, the credentialed and the uncredentialed, and the sincere and the deceptive. Despite all of these similarities, the most popular guitar tutorial content creators typically have significant expertise at guitar, but a significant portion of the most popular medical content creators do not have sufficient expertise and/or often mislead users.

In what follows, we solve this puzzle by rethinking traditional ways of framing the problem of recognizing experts. Philosophers concerned with this problem have primarily focused on whether (and when) *individuals* are capable of (or likely to) recognize experts (e.g. Goldman, 2001; Anderson 2011; Guerrero 2016; Levy, 2019; McKenna, 2023). Solutions have thus often focused on improving the likelihood that individuals recognize experts, whether by education (either teaching information or skills (e.g. Caulfield & Wineburg 2023)), attempts to reduce bias,² or improvements to the overall epistemic environment (e.g. interventions on credentialing institutions, media, social networks, etc. (Anderson 2011; Guerrero 2016; Levy, 2019; Marin, 2022; Amico-Korby et al., 2024)). We are friendly to many of these proposed solutions, but we should not conclude that, if they fail, the problem of recognizing experts is intractable. Rather, we will demonstrate that the *crowd* can

¹ One interesting exception is how YouTube handled medical content after the spread of Covid-19 misinformation on the platform (*Get info on health-related content—YouTube Help*; Currin 2021).

² See Kovaka (2021) for discussion of these first two approaches.

recognize experts in certain conditions, even when individuals would struggle to recognize them.

More specifically, we claim that social media platforms have enabled a novel wisdom of the crowd phenomenon to emerge. Specifically, the crowd—in combination with search and recommendation algorithms—can reliably identify DIY experts and thus fulfill one of the main functions of a credentialing institution. However, we also claim that a necessary condition for this wisdom of the crowds phenomenon to get off the ground is that individuals are (at least) somewhat reliable evaluators of individual pieces of content.³ This condition fails to hold, we argue, for many types of non-DIY content, including content about climate science, vaccine safety, and election safety.⁴ Thus, the crowd—in combination with search and recommendation algorithms—is unable to serve as a reliable credentialing institution for non-DIY content creators. That is, evaluations of DIY content by individuals are reliable enough that when aggregated the crowd reliably picks out experts, but this doesn't hold for non-DIY content.⁵

Importantly, while our focus will be on defending and explaining one instance where the crowd can recognize experts (DIY content experts), we will provide a more general recipe for determining when (and how) the crowd can succeed at this task. Specifically, we will provide a set of conditions for when we should expect individuals to be reliable (enough) judges of individual pieces of content, and suggest that, in those conditions, appropriate aggregation methods should result in a wise crowd.

This solution is particularly important because many authors have noted individuals' difficulties in recognizing experts (Goldman, 2001; Guerrero 2016; Levy, 2019). One of the primary difficulties is the vast number of domains in which we all must rely on experts, including bridge building, virology, plumbing, food safety and cooking, automobile engineering, water treatment, weather prediction, and so many more (Milgram 2015). With sufficient time, most people could presumably recognize experts in each of these domains. Indeed, with sufficient time, most people could become experts themselves in most of these domains. But none of us can do this work for *all* of these domains. The most common social solution to this problem is credentialing institutions. But such institutions are costly to create, maintain, and oversee. Our solution shows that, for some subject matters, there is another solution. The crowd can recognize experts and serve as a credentialing institution, even when no individual would be in a position to do this work themselves.

Our proposal may seem surprising because the wisdom of the crowds is often thought to contrast with the wisdom of experts. For example, scholars have typically focused on determining when the crowd is as reliable (or more so) than experts,

³ As we'll point out, there are many cases where one might be a somewhat reliable evaluator of individual pieces of content but still struggle (or fail to be in a position to) recognize experts.

⁴ While there are many types of non-DIY content, for ease we will use "non-DIY content" to refer to informational non-DIY content throughout.

⁵ The wisdom of the crowds has been explicitly connected with search and ranking algorithms at least since Brin & Page's (1998) influential PageRank paper. For an overview of more current work on the wisdom of the crowds and PageRank see Masterton & Olsson (2018).

so that they might determine when it is reasonable to replace decision making by experts with decision-making by the crowd. It is thus a significant upshot of our proposal that we can sometimes harness the wisdom of the crowds to recognize experts.⁶

Our proposal may also seem surprising in the context of recent (predominantly negative) academic and non-academic discussions of search and recommendation algorithms, as well as social media platforms more broadly. Platforms have been accused of shortening attention spans (Hari, 2022), contributing to the spread of misinformation (Barrett & Hendrix, 2022; Myers, 2022) and the formation of filter bubbles and echo chambers (Sunstein, 2017; Nguyen 2020a), and increasing polarization (Barrett et al., 2021). Recommendation algorithms specifically have been accused of guiding our attention towards the wrong things (Schuster & Lazar, 2024), making YouTube “the great radicalizer” (Tufekci, 2018), taking us down the rabbit hole (Alfano et al., 2021; Tang et al., 2022), of being “the Great Polarizer” (DiResta, 2018), and—by an ex-YouTube designer—for increasing the glut of misinformation on YouTube and allowing “fiction to outperform reality” (Lewis, 2018).⁷ Some have viewed things as dire enough to call for radical changes to social media. For example, Regina Rini has suggested that we should perhaps “blow it all up”:

If we can convince people to become universally skeptical of social media content—to disbelieve everything conveyed through the medium [...] then we can at least blunt the effects of induced turmoil [...]. On this strategy, social media needs to be brought down to the epistemic level of children’s cartoons; a source of diversion, but not a place any adult would think to go for information (Rini, 2021).⁸

Importantly, these concerns are not simply academic. For example, New York (U.S.) Governor Kathy Hochul has recently signed a state bill (the Stop Addictive Feeds Exploitation (SAFE) for Kids Act) that prohibits social media platforms from serving content to users under the age of 18 based on recommendation algorithms (Tenbarge, 2024). Other states are likely to follow suit. In the EU, the Digital Services Act places constraints on the recommendation and content moderation systems used on large social media platforms. More broadly, governments around the world are grappling with regulation of recommendation algorithms, and for good reason. People increasingly get their news and information from social media platforms (54% of US adults, per Pew Research (Social Media and News Fact Sheet)), and “teenagers swim in a sea of digital information.” (Breakstone et al., 2021) More than 75% of students in the US report getting their news “often” or “sometimes” from social media, and 69% report getting their news from videos on sites like YouTube

⁶ In a sense, this still constitutes a replacement of expertise. After all, by recognizing expertise the crowd is fulfilling an important function of credentialing institutions, and the work of such institutions is often performed by experts. For example, medical school curriculum and certification exams are typically designed and created by medical experts.

⁷ Recent updates to YouTube’s algorithm appear to have mitigated some of the concerns about radicalization (Brown et al., 2022).

⁸ See also Campbell (2024) who argues that we have a duty to quit using social media platforms.

(Dautrich, 2018). Crucially, people (especially young people) do not merely *see* news on social media platforms, but often believe and trust it. According to Pew research, U.S. adults under 30 indicate that they trust information they get from social media sites almost as much as national news outlets (Trust in Media).

In this context, we think it is especially important to draw attention to this wisdom of the crowds phenomenon. It is a significant and overlooked epistemic good that social media platforms—and recommendation algorithms specifically—have helped provide, which is worth understanding for its own sake (how is it that an algorithm that sorts content based on predicted engagement reliably produces high quality epistemic results?). Moreover, in the context of broader political conversations about how to regulate social media, it is essential to understand the benefits of social media alongside its risks and costs. Further, it is essential to understand the epistemology of these algorithms for two reasons. First, people increasingly depend on algorithms to present their news and information, and so we should analyze when it is reasonable to do so, in order to provide appropriate advice and education to social media users. Second, platforms and democratic governments can only make intelligent decisions about recommender systems if we properly understand them. Blowing it all up, as Rini suggests, may ultimately be what is required, but doing so would come at great cost.

Our proposal implies that the wisdom of the crowds phenomenon does not work for all content. For this reason, our final section proposes a concrete strategy for platforms to implement when the wisdom of the crowds fails. While we do not argue that our proposal should be a legal requirement, we contend that platforms must make some adjustments when crowds fail to identify experts. We propose that they should rely on real-world credentialing institutions with democratic oversight to help them sort their content. Importantly, our proposed strategy (1) doesn't require platforms to make epistemically fraught decisions, (2) involves democratic oversight, and (3) is easy for platforms to implement.

We proceed as follows. First, we identify and explain this novel wisdom of the crowds phenomenon. Specifically, we argue that the crowd can reliably pick out experts and serve as a credentialing institution. Second, we defend our explanation for the puzzle just introduced. We argue that (a) five epistemic features determine whether the crowd can succeed at recognizing experts on social media platforms, and (b) while many of those features are satisfied to a sufficiently high degree by DIY content, they are mostly lacking for non-DIY content (e.g., about climate change or vaccines). Third, we situate our explanation in the broader literature on explanations for why people form problematic beliefs. And finally, we argue that our analysis of the puzzle implies that social media platforms ought to treat content that satisfies our factors and content that does not differently.

2 The Phenomenon

Generally, a crowd is wise whenever it produces accurate judgements (Thorn & Schurz, 2012). In traditional examples, authors have often measured accuracy by comparing the crowd's accuracy to that of expert individuals. The best-known

examples of the wisdom of the crowds effect are for estimates of physical quantities (e.g., how much a cow weighs, how many jellybeans are in a jar, or how many people live in a particular country). In these cases, an aggregate of individual people's judgments is often more accurate than almost any individual judgment, though this can depend on the specific method of judgment aggregation (Galton 1907; Surowiecki, 2005; Solomon, 2006; Page, 2007; Keck and Tang 2020).

We contend that social media platforms have enabled a novel wisdom of the crowds phenomenon to emerge. Specifically, the crowd—in combination with search and recommendation algorithms—reliably picks out DIY experts much better than individuals can, and thus fulfills one of the main functions of a credentialing institution. Here's how this works. Individuals are somewhat reliable evaluators of individual pieces of DIY content on social media platforms. When people positively evaluate DIY content, they often react in measurable ways (e.g. liking, sharing, subscribing to the channel, or continuing to consume content from the creator). When people negatively evaluate DIY content, they are unlikely to continue consuming content from that creator or react to it positively (e.g. by liking or subscribing). That is, people's epistemic assessments of DIY content are a significant factor in their engagement patterns for that content.⁹

Of course, these individual epistemic assessments are fallible and limited. First, the epistemic quality of content is only part of what drives people's engagement decisions. For example, people may not react positively to high-quality information if it is about something they find gross or disgusting. Similarly, people may find bad content amusing in certain contexts (perhaps someone whose cakes always turn out hilariously bad) and react positively to it.¹⁰ Second, most people only assess a minuscule sample of content for any domain, and thus have limited evidence about the overwhelming amounts of content on platforms (e.g., YouTube has ~ 14 billion videos per McGrady 2023). Third, people spend very little time on any piece of content (Nielsen, 2011). Fourth, while people are somewhat reliable evaluators of individual pieces of DIY content, many individuals lack significant expertise in the DIY content they consume.

These four limiting factors imply that many individuals are unlikely to be capable of doing what the crowd, in combination with the relevant algorithms, does.¹¹ For example, suppose individuals were given access to YouTube's 14 billion videos sorted by upload date. Even if they were highly reliable at evaluating individual pieces of content in the many DIY subject matters represented on YouTube, it would take an extraordinary amount of time to identify and rank subject matter experts. And the average person probably is not highly reliable (e.g., they are likely to miss

⁹ Recommendation algorithms from their inception were designed to exploit the wisdom of the crowds, but not for purely epistemic ends. Information was too abundant and too overwhelming online, and recommendation algorithms were proposed as a way of harnessing the crowd to determine what individual users were likely to find engaging and useful (cf. Shardanand & Maes' 1995 classic paper). Given this, it is somewhat surprising that less attention has been paid to when and how these algorithms can be used to produce epistemically good results.

¹⁰ We discuss this in greater detail when discussing factor 3 below.

¹¹ Of course, there will be some members of the public who have significant expertise in various DIY subject matters.

nuance that an expert would detect). Despite this, as in more traditional wisdom of the crowds phenomena, fallible and limited individual assessments can lead to highly accurate group assessments when appropriately aggregated.

Successful aggregation depends on the search and recommendation algorithms. The technical details of these algorithms are unnecessary for our purposes, but at a conceptual level these algorithms preferentially display content according to engagement metrics such as likes, shares, subscriptions, views, and watch time, perhaps weighting reactions from individuals like the user (in some sense).¹² Search algorithms sort items in search (Rieder et al., 2018), and recommendation algorithms shape the display and order of content on home pages, feeds, and up-next panels (Goodrow 2024).¹³ Our contention is that these algorithms use the behavioral signals generated by individuals' (somewhat) reliable epistemic assessments of the quality of content, and then output lists of DIY content creators who *in fact* are experts in their respective areas.

Of course, these algorithms aggregate individual judgments very differently than in traditional instances where the crowd is wise. In traditional instances, people make judgments about a question (e.g., “size of p ?”), and then those judgments are aggregated (e.g., compute the mean) and compared with the accuracy of those individual judgments (especially those made by experts). Social media algorithms, in contrast, aggregate behavioral indicators (e.g., engagement) rather than explicit judgments, and their main output answers a question that most individuals were not attempting to answer. Individuals assess the quality of individual pieces of content, and the algorithms output a list of DIY experts.

There are two important upshots of this phenomenon. First, these algorithms produce aggregate judgments of DIY content that improve accuracy and provide accurate assessments of questions that individuals didn't directly try to answer. In fact, they may even be sufficient to act as (informal) credentials. Credentialing institutions are extremely costly (in money, time, and people) to design and run. Given that these systems are capable of accurately identifying DIY experts, they may be able to supplement or replace these institutions in some circumstances.

Second, these algorithms enable individuals to rapidly identify high-quality content and high-quality experts to learn from. For most of human history, if one wished to learn (or improve at) a skill, one had to either make progress on their own (reinvent the wheel, so to speak) or do the work of finding an expert to teach them (or some combination of both). And the work of finding experts to trust is arduous, especially for novices.¹⁴ Credentialing institutions have made the task of finding experts much easier, as one can rely on credentials to identify who to

¹² Relevance to the query is also an important factor for search algorithms.

¹³ Different platforms prioritize different engagement metrics (Narayanan, 2023; Thorburn, 2022). We largely focus on YouTube because of its importance as a platform for DIY content (and as the second most visited platform worldwide as of November 2022; McGrady et al., 2023), for which watch time has long been one of the most important factors (Meyerson 2012; Covington et al., 2016; *A Few Factors in Search Ranking—YouTube Help*).

¹⁴ See, for example Goldman's (2001) discussion of determining the quality of putative experts without relying on credentials, and Guerrero's (2016) discussion of the difficulty of determining who to trust even when one has credentials.

trust—perhaps a licensed plumber or a Julliard-trained guitarist. But there are three problems with this approach. First, while it’s somewhat easier to find these experts, learning from them remains expensive. Thus, on the traditional model, many people are constrained (for monetary reasons) in how many skills they can study and develop. Second, experts have limited time. For instance, individual guitar experts can only teach so many students in-person at the same time, thereby increasing costs and reducing learning opportunities. Finally, in the context of social media, credentials (or markers thereof) have not been built into the architecture of platforms. One cannot, for instance, search YouTube for “Julliard-trained guitarists” or “licensed plumbers” and find a list of videos created only by those who meet those qualifications. The lack of credentials contributed to the character of social media platforms in their early days, where scrolling YouTube was an adventure filled with videos of extremely low quality and unhelpfulness, but also the occasional gem.

The wisdom of the crowds phenomenon we’ve identified plays a significant role in addressing each of these problems.¹⁵ People now have cheap and easy access to high quality instruction in an extremely wide range of domains (though not all!). It is now easy to find high-quality instructors and high-quality information because of the wisdom of the crowds phenomenon, even without platforms integrating institutional markers of credentials into their design. People are thus able to pursue their ends and develop their perfections (Brink 2013, 2019; Bradford 2017) in a way and to a degree only previously available to the extremely rich and privileged. This, we submit, is a remarkable and important epistemic and moral benefit that social media platforms have provided.

At the same time, as previously noted, this phenomenon does not always work. Sometimes, the crowd does *not* find useful experts. The next section explains and defends five factors that influence people’s epistemic reliability (and behavioral indicators) for individual pieces of online content (i.e., when the wisdom of the crowd phenomenon will be able to get off the ground). We do not attempt to answer precisely how reliable individuals must be, not least because *there has been (to our knowledge) no empirical (or philosophical) work on this phenomenon*. That being said, we believe that this is the sort of phenomenon whose mere description has the ring of truth to those who are familiar with it. For those with less familiarity with social media platforms, we encourage you to open YouTube (or other another social media platform with a robust DIY community) and search the DIY subject matter of your choice. We predict you will find that for most DIY subject matters, the people who have the most followers and the most consumed content typically have a high degree of expertise. For empirical researchers, we are hopeful that our proposal will seem plausible and important enough to be worth examining in greater detail.

Nonetheless, we now must consider seriously why our proposal does not also apply to content about, say, vaccines or climate change. It would be surprising to find popular content creators peddling bad cake recipes (in a non-comedic manner),

¹⁵ An interesting question that we do not answer here, is whether (and which) alternative aggregation methods for SM platforms would result in a similar phenomenon. We’re grateful to Phil Chodrow for discussions of this point.

but content creators who peddle election, vaccine, and climate misinformation have all (to various degrees and at various times) become very popular.¹⁶ Alex Jones, before being deplatformed, had 2.4 million YouTube subscribers (Coaston, 2018; Newton, 2018), and his site InfoWars had 10 million unique visits in a month at one point, comparable to *Newsweek* and *The Economist* (Beauchamp, 2016). Robert Malone (noted vaccine skeptic; Alba, 2022; Bartlett, 2021; Bella, 2022) has 1.1 million followers on X; Dinesh D’Souza (a spreader of election and Covid-19 vaccine misinformation; Bump, 2023; Reuters Fact Check, 2022) has 3.6 million followers on X and 750 thousand subscribers on YouTube; and Dr. John Campbell (a YouTube account that popularized Covid-19 misinformation; Gorski, 2022; Cercone 2024) has 2.88 million subscribers. Deceptive accounts created by the Russian Internet Research Agency before the 2016 presidential election collectively gained hundreds of thousands of followers (US v. Internet Research Agency 2018). Many of these followings were built despite various interventions from platforms (e.g. removal of content and downgrading in search (*Get info on health-related content—YouTube Help*)). These unreliable content creators all managed to generate comparable or more engagement than reliable medical content creators like the CDC (5.5 million X followers, and 628 k YouTube subscribers) and The Mayo Clinic (2 million X followers and 1.01 million YouTube subscribers).¹⁷ Content creators who mislead have, in many cases, amassed comparable followings to those who are trustworthy, and people do not seem to consume this content idly or solely for entertainment purposes. Polling data indicates that: many Americans (estimates range between 32 and 52%) do not believe in human-caused climate change (Funk and Kennedy 2016); 36% of Americans believe the risks of the Covid-19 vaccine outweigh the benefits (Funk et al., 2023); and three-quarters of Trump voters incorrectly believe he won the 2020 election (Pew Research Center 2020).¹⁸ Finally, we know from extensive research that much medical information on YouTube is not of high epistemic quality (Mamo et al., 2021; Crutchfield et al., 2021; Etzel et al., 2022; Tang et al., 2022; Ayo-Ajibola et al., 2024),¹⁹ as videos created by physicians often do not receive more views and/or likes than content made by non-experts (Celik et al., 2020; Esen Özdemir et al., 2023; Ayo-Ajibola et al., 2024), and higher epistemic quality of videos is often not associated with increased viewer interaction (Birch et al., 2022; Celik et al., 2020; Murugiah et al., 2011).²⁰ Thus, we know that the crowd is not

¹⁶ We focus on American examples due simply to our knowledge of them, not because this phenomenon is unique to the US.

¹⁷ These sites are some of the most popular, and most clearly trustworthy, though they are not the only such sites. See also Jaworska et al., 2024 for more on what they call the “Twitter influencing elite” posted about Covid-19, where they find thatrealDonaldTrump and ElonMusk were two of the top 6 most retweeted and favorited accounts for Covid-19 messages, both of which often spread problematic content.

¹⁸ There is also some evidence that exposure to misinformation decreases belief accuracy (Porter et al., 2023).

¹⁹ In the worst cases, the only information on a topic is made by non-experts, is extremely inaccurate, but nonetheless appears to be positively received by consumers (Mamo et al., 2021).

²⁰ There are some unusual cases. For example, high quality videos on Transcranial Magnetic Stimulation in stroke both generate the most views, but also more dislikes than other videos (Askin et al., 2020).

wise for these particular topics, but we do not yet have an explanation why.²¹ We now turn our attention to solving this puzzle.

3 Five Factors that Determine when it's Safe to Trust the Crowd

We contend that five epistemic factors—framed here as questions—determine whether we should trust the crowd to recognize experts in a particular subject matter. Each factor comes in degrees, and the higher the degree of satisfaction across all factors collectively, the more likely it is that it would be reasonable to trust crowds to identify experts in that subject matter. We do not provide precise thresholds (e.g. when these factors are sufficiently satisfied to warrant trust in the crowd), as different subject matters will satisfy these factors in different ways, and the importance of the various factors may vary by subject matter. Nonetheless it will always be the case that a higher degree of satisfaction for a factor increases the likelihood that the crowd is wise, and that it is thereby reasonable to trust the crowd.

The first two factors are closely related to what Goldman (2001) and Guerrero (2016) call ‘esoteric’ and ‘exoteric’ statements. According to Goldman, ‘esoteric’ statements are those that an expert makes where the truth values are inaccessible to novices. Truth values of ‘exoteric’ statements, in contrast, are accessible. Our first two factors can be understood as identifying ways that content can vary in its degree of accessibility to the public, and so are related to the exoteric-ness of the content. We emphasize, however, that we are applying these concepts to a different task than these authors. Goldman, for example, is interested in when a lone individual could determine which of two putative experts they should trust. In contrast, we use these concepts to help identify the conditions under which the *crowd* can recognize experts.

A key background assumption is that, while most people consume large amounts of online content, they spend very little time on individual pieces of content. Hence, successful instances of the wisdom of the crowds phenomenon will often depend on people either being able to quickly determine whether the content is likely to be good, or having incentives to invest the time to test the advice (e.g. baking a cake). Moreover, we are interested in the crowd’s ability to serve as a credentialing institution, and so we assume that the public can (at least, sometimes) judge expertise without a credentialing institution. With those background observations in place, we turn to the specific factors.

Factor 1 *Does the public know what evidence is relevant to the evaluation of purported answers, solutions, or performances?*

For many DIY cases, this is so clear that it almost sounds silly to point it out. It is obvious that we evaluate cakes based on taste, texture, and looks. Evaluating a performance of a song involves evaluating how the song is played. And so on. Of

²¹ See also Sullivan et al. (2020) who apply Google’s PageRank to Tweets about vaccine information to see if this generates a wisdom of the crowds phenomenon and found that it did not.

course, we typically can't evaluate the taste and texture of a cake made by the DIY content creator. Thus, we must evaluate some of these features indirectly, which is usually relatively straightforward. For example, we might attempt to follow the putative expert's directions precisely, and then (a) check the cake's taste and texture given (b) how closely we were able to mimic the instructions and recipe that were provided. If the cake turns out poorly, and we're reasonably confident that we mirrored their directions, then we have a clear indication that the instructor who made the video is not an expert (or perhaps is trying to prank us or lead us astray).²² The evidence to evaluate in these cases is straightforward and clear.

However, other cases, particularly outside of the DIY realm, are not like this. For example, imagine that you're invited to watch a sport that is new to you, and are given the task of evaluating player performance. You would likely find this task difficult, as you do not know the objective of the game, the rules of the game, the roles and assignments of various players, and so on. This information is crucial for understanding what evidence matters, how to interpret and think about your evidence, and so on. Thus, to the extent that someone lacks this information, they will struggle to accurately evaluate the performance.²³

We contend that topics such as global warming are more similar to novel sports than cake baking. What should we evaluate if we are trying to determine whether the climate is warming? What evidence matters for answering this question? What are the rules governing what evidence matters and how to produce or evaluate it? Most of the public is unable to immediately answer these questions in an informative way. The following comment made by a person in the US (who sees little urgency to address climate change) reveals a lack of understanding about what to evaluate:

I just don't think that there's any evidence. I mean I don't see where anything has changed throughout my life. The summers aren't hotter. It's just not any different ~ Woman, 40 s, Mountain West (Pasquini 2023).

Or consider former US senator James Inhofe who brought a large snowball to the US Senate floor, apparently as evidence that it was unseasonably cold and thus concerns about climate change were overstated (Barrett 2015). In both cases, these individuals appeared to not know *what* to evaluate in order to determine whether the climate is warming; they focused on evidence regarding (spatially and temporally) local weather as opposed to evidence about global climate.

There are many other examples in which the public typically does not know what to evaluate in order to answer a question. Imagine being asked to determine (without relying on credentials) whether someone was qualified to analyze data from the Large Hadron Collider. We should not expect individual members of the public to have any idea what evidence is relevant to making such an assessment.

The climate change and Large Hadron Collider cases mirror the novel sport case in several ways, but not all ways. To the extent that one lacks a clear understanding

²² There will also be cases where it's difficult for non-experts to determine whether they've appropriately followed the directions.

²³ Of course, possession of this information does not require conscious awareness or even conscious accessibility.

of the objectives and rules of inquiry, of performance, etc., one will find it difficult to know what evidence matters, what to evaluate, and how to evaluate it. There is, however, an important difference between these cases in terms of *accessibility* of information about what evidence matters. For new sports, the relevant information is typically easily accessible at little cost to non-experts. The rules are often publicly available, friends might share information, commentators will provide information, and so on. But in the case of climate science or particle physics, knowing what evidence matters will often require an immense amount of domain-specific knowledge and education. And that is the sort of thing that is not readily possessed nor easily attainable by the average person without significant time and effort. Most importantly for our purposes, it's not the sort of thing that most people take the time to develop before they make rapid evaluations of content on social media platforms.

Straightforwardly, if the public doesn't know what evidence is relevant to answering a question, we shouldn't expect them to be reliable enough judges of answers to that question. To the degree that important questions in a field often have this feature, then we should not trust the crowd to evaluate expertise. We can thus capture the lesson of Factor 1 as follows: The more difficult it is for non-experts to know what they are supposed to evaluate, the less likely it is that they'll be adequate judges of purported answers.

Factor 2 *How easy is it for the public to evaluate purported answers, solutions, or performances?*

A closely related way of putting this factor is in terms of the public's degree of skill or expertise at evaluating content in the relevant area. This factor is distinct from the first, as it is one thing to know what evidence is relevant to the evaluation of purported answers, but a very different thing for it to be easy to actually evaluate them. For example, when I watch grandmasters play chess, I know that I should evaluate their positioning, the protection of their king, their pawn structure, and so on. However, doing this well is extremely difficult. Most of the general public lack significant expertise at this task.

In general, answers, solutions, and performances can vary in how easy they are for the public to evaluate or in terms of what proportion of the public have significant skill in making evaluations. In the best cases, (nearly) everyone is capable of evaluating proposed answers. For example, if the question is "what are today's specials on the menu?", then everyone who looks at the menu (and speaks the language of the menu, has normal vision, etc.) will be able to answer the question and evaluate others' answers to the question. In many DIY cases, the relevant evaluations are relatively straightforward. For example, it is easy to determine whether one's sink is still clogged by checking whether water is pooling in it. In other cases, the average member of the public may find evaluation initially difficult but could master it with little time and effort. For example, someone who has never had alcohol or spicy food might struggle to evaluate the quality of a whiskey or a papaya salad due to the strong flavors of whiskey or the intensity of the spice. Even if this person knew what sort of qualities they should be trying to evaluate, they might be unable to detect

them. However, with relatively minimal effort, most people can build up sufficient tolerance and experience to make better evaluations of whiskey and spicy foods.

In the worst cases, however, it is very difficult for non-experts to evaluate purported answers correctly. The public is unlikely to have the skill to evaluate the quality of a proposed proof of Fermat's last theorem, the quality of a climate modeling technique, or the quality of a new jet engine prototype, even if they know the types of evidence that would be relevant. And they certainly won't be able to do so in the timeframes of social media content consumption.²⁴ As a result, we should expect the wisdom of the crowd phenomenon to fail when people cannot quickly and reliably (enough) evaluate the relevant subject matters (even if they know what evidence to consider).

Although evaluation difficulty usually impairs assessment of expertise, there may be some exceptions. In particular, we must distinguish between direct and indirect evaluation. Consider a case (adapted from Nguyen, 2020b, p. 2806):

Meteorology: Noah is not a meteorologist. He knows very little about the science of weather prediction and has very little idea about how to answer questions in meteorology. Thus, when a purported expert meteorologist makes a weather prediction, before the date of the prediction, Noah is not able to evaluate the prediction's accuracy nor plausibility. Nonetheless, once the date of the prediction arrives, Noah can evaluate the accuracy of the prediction.

In this case, Noah cannot evaluate the meteorologist's predictions *ex ante* in the way a fellow meteorologist would. Still, can get evidence regarding the accuracy of their predictions by checking the weather on the corresponding day. That is, the evaluation of expertise can proceed by confirmation of specific predictions.

Before moving onto factor 3, consider an example to underline how failures of factors 1 and 2 show up in real life, even for the highly educated. Sam Wineburg & Sarah McGrew (2019) asked small groups of Stanford undergraduates and Stanford history professors to evaluate the credibility and plausibility of various pieces of content on the web. Participants were provided the content, but also a full web browser and the freedom to use the wider internet in their task if they wished. Surprisingly, only two out of ten of the Stanford historians evaluated content adroitly; they were often no better than undergraduates at making evaluations. All participants sometimes rated problematic materials produced by hate groups as having high credibility and plausibility. The explanation for these failings, according to Wineburg & McGrew, is that both historians and undergraduates tended to rely on their own ability to make evaluations when examining content. When presented with a web page, they would scroll the page, read the content, perhaps check the "about" page on the website, and then attempt to evaluate the article based on its own merits and whatever indicators of professionalism (like citations or professional seals) or

²⁴ Consider the following comment about climate change made by a person in the US: "I think there's been extreme weather events since the beginning of time. ... There's been hurricanes, tornados, all that. I just don't buy into the fact that they're getting worse" –Man, 50 s, Coastal Florida (Pasquini 2023). This individual seems to know one type of evidence that matters for evaluating the question (factor 1), but struggles to evaluate, perhaps because determining the frequency of severe weather events is difficult using only personal experience.

accuracy they could find on the page. However, because they lacked expertise in the relevant areas, even very smart members of the public (Stanford historians!) were easily taken in by deceptive and problematic content and made many mistakes.

The former study involves small samples, but Breakstone et al. (2021) echoes the underlying lessons of that paper in survey work with 3,466 high school students. When the students were asked to investigate a site claiming to “disseminate factual reports” on climate science, 96 percent never learned about the organization’s ties to the fossil fuel industry. More than half believed that an anonymously posted Facebook video, shot in Russia, provided “strong evidence” of US voter fraud. Much like the Stanford historians and undergrads, one way these high school students seemed to run into trouble was by evaluating the content on its own merits or on the merits it claimed for itself. For example, in the climate science case some students cited factors like the website having a “.org” at the end of its URL as evidence for its reliability, or that “they [the website pages] provide evidence for both sides and state matters in a scientific manner”. Regarding the video shot in Russia, many students simply took the video at face value: “Yes, it shows video evidence of fraud in multiple different states at multiple different times”, and “[it provides strong evidence of Democratic voter fraud] because the video showed people entering fake votes into boxes.” These examples suggest that the participants often did not know what evidence was relevant for making accurate assessments and lacked the skill to make the assessments when they did.²⁵ Of course, even when people know what to evaluate and have the skill to evaluate it, they may not have the time to make such evaluations.

Factor 3 *What is the time frame for evaluation?*

People rarely spend significant time on any particular piece of social media content (Nielsen, 2011), even though some content requires significant time to evaluate. For example, suppose a climate scientist predicts that global warming will increase the frequency of severe weather over the next 50 years. Direct evaluation is unlikely to be successful (due to the previous two factors), and people cannot wait for 50 years to determine accuracy through indirect means. In general, the longer the time frame for evaluation, the less likely it is that the crowd will be able to successfully recognize experts.

People sometimes have epistemic or non-epistemic reasons to invest time in testing or evaluating the content. For example, people may spend considerable time to improve their evaluation of DIY content, particularly if it improves their ability to learn a skill or accomplish a task. People wishing to learn how to play guitar, bake, better understand their car engines, or many other (but not all) activities will need to make significant time investments. Crucially, the time investment will frequently put them in a better position to evaluate the content quality. For example, imagine one watches multiple videos from the same guitar instructor over the course of several

²⁵ See also Caulfield & Wineburg (2023) for extensive discussion of these failings as well as practical lessons for doing better. Kreps & Kriner (2020) provide further evidence of people’s failures to distinguish fact from fiction, this time with respect to Covid-19.

weeks and spends the requisite time practicing what was taught. By the end of this period, they should be able to recognize improvement—they can play a new song, they’ve learned a new scale, they’re developing a new technique. The time spent following the DIY advice thereby improves one’s ability to evaluate the content.

Of course, this doesn’t always hold. For example, people may consume DIY content purely for entertainment, as poor DIY content (bad cooking instructions, for example) can be particularly amusing. Nonetheless, we do not think this seriously undermines the wisdom of the crowds phenomenon for two reasons. First, when one searches (for example) for “vanilla cake recipes” on YouTube, humorous “bad” DIY videos are not at the top of the recommended videos. Second, we hypothesize that this type of content tends to occur in those areas where the average member of the public has the most expertise and is thus the least likely to be led astray by bad content. The humor of such content often requires that the audience be in on the joke. TV shows like “Nailed It,” where amateur bakers with poor baking skills often fail at difficult challenges, wouldn’t be as funny if the audience wasn’t aware that they were failing.²⁶

In contrast, non-DIY content often does not incentivize spending extra time with the content to self-improve or otherwise learn to evaluate the content. For example, the “news-finds-me” effect—the perception some individuals have that they are well informed without intentionally following the news because the news finds them (e.g., through peers and social media)—is well documented (Gil de Zúñiga et al., 2017; Park & Kaye, 2021). In these cases, rather than actively seeking out and practically testing information, people wait on information to come to them. When this effect is active, we should expect people to spend limited time on each piece of content, and people’s evaluation of that content will likely be correspondingly poor.²⁷

Relatedly, one might wonder whether individuals who consume news and non-DIY informational content on platforms form an epistemically problematic subset of the population. Perhaps, for example, those who wish to find reliable information rely on traditional news sources or reliable institutions like the CDC, while the less epistemically virtuous get news and information from social media platforms like YouTube. In such a case, engagement markers of news and informational content on social media platforms would be distorted by the skewed population, and the wisdom of crowds phenomenon would likely fail.²⁸

Without extensive empirical work, it is difficult to evaluate the likelihood of this scenario, but we note several things that suggest it should not be overstated. First, as we previously observed, significant portions of the population get their news and information from social media platforms. Second, many reliable news (and other) organizations and experts spend considerable time producing social media content. Until very recently (and arguably even still), X (formerly known as Twitter) had an extensive journalistic, academic, and “expert” presence. Moreover, reliable sources on social media platforms are often very popular, though not necessarily

²⁶ We are grateful to an anonymous reviewer for encouraging us to discuss this point.

²⁷ A similar phenomenon could occur if people consume content to affirm their identities. We briefly discuss this point in the next section in the context of non-DIY content.

²⁸ We are grateful to an anonymous referee for encouraging us to explicitly address this.

more popular than unreliable sources.²⁹ Finally, even if something like this hypothesis holds today, we are skeptical that it will continue to hold in the future. A survey of nearly 10,000 high school students in the US found that over 75% report getting news from social media sometimes or often, and especially from videos on websites like YouTube (Dautrich, 2018). Social media has become the dominant form of content on the internet, and it is likely that going forward most people (the epistemically virtuous and vicious alike) will get at least some news and information from these sites.

Let us reiterate the central point of this section. Successful instances of the wisdom of the crowds phenomenon depend on people either being able to quickly determine whether the content is likely to be good, or having incentives to invest the time to test the advice.

The previous three factors focused on constraints on the assessor, but the reasons we require expertise are also relevant.

Factor 4 *How fine-grained are the distinctions between degrees of expertise that we need to make?*

Factors 1–3 come in degrees, and they may be satisfied to a high degree for some questions in a subject matter and not for others. Consider DIY fitness. Some aspects will be straightforward for people to evaluate, such as changes in their performance or physical appearance over time. But it will also involve some elements that people are probably not well prepared to evaluate, such as the long-term health effects of certain supplements or diets. In cases like DIY fitness, the crowd might reliably pick out experts in a subset of a subject matter (e.g., those with expertise about improving athletic performance), but not for other parts of the subject matter (e.g., the long-term health effects of supplements or diets). The wisdom of the crowds will only be broadly useful if expertise in the former subset predicts expertise in the latter.

Relatedly, we should expect the crowd on a platform like YouTube to be fairly lenient about who counts as an expert, even in those cases where the wisdom of the crowds phenomenon is successful. Because of this, when the stakes are high, we may wish to rely on a more traditional credentialing institutions to pick out a narrow range of experts with sufficiently high degrees of expertise.

There are two reasons why we should expect this. The first is that the incentives and stakes that individual evaluators have when evaluating a piece of content vary widely. If I wish to learn guitar basics, I do not need to identify one of the best guitar players or even one of the best guitar teachers in the world. Many mediocre players and teachers could show me how to fret basic chords and help me accomplish my goals. Thus, when low degrees of expertise will satisfy the needs of a sufficiently large proportion of a subject matters' content consumers, we should expect the crowd to pick out a fairly wide range of people as suitable experts.

Second, in some cases, more precise distinctions between degrees of expertise will require that the crowd do better on factors 1–3. For example, it may be that the

²⁹ Another way of putting this is that the crowd in combination with recommendation algorithms doesn't distinguish between experts and non-experts.

crowd is perfectly capable of picking out pretty good pianists. The public could certainly distinguish between the authors' ragged and halting performances of Chopin's Nocturnes and a virtuoso's performance. However, many members of the public will not have an awareness of the range of techniques required to properly perform Chopin's Nocturnes, nor the ear to recognize whether a particular pianist has executed them.³⁰ As the crowd does worse with respect to factors 1–3, we should expect the crowd to include a wider and wider range of people in their pool of experts.

One may wonder whether this matters. Lay people should rely on expert consensus, but if there is no expert consensus, then individual experts' opinions shouldn't be accorded special weight. What matters is whether there is consensus amongst experts, and whether those we learn from and listen to reflect that consensus.³¹

We agree that lay people often should rely on expert consensus when they can. Nonetheless, we have identified two distinct methods of determining who the experts are: traditional credentialing institutions and the wisdom of the crowds. Answering the question "is there expert consensus?" will (in some cases) depend on which method we use because the methods are likely to differ in how stringent they are about who counts as an expert. For example, there may be some questions that all physicists with very high degrees of expertise agree on, but that those with some (but not a very high) degree of expertise lack the skill to evaluate and often get wrong. Including a wider range of people as experts in a case like this may result in the appearance that there is no expert consensus on these questions. In a case like this, it would be better to rely on the consensus of experts as credentialed by traditional credentialing institutions than as credentialed by the crowd. Thus, while we agree that lay people often should defer to expert consensus, when the stakes are high, we may wish to rely on traditional credentialing institutions to pick out who the relevant experts are.

Factor 5 *Is the crowd sufficiently epistemically independent?*

Finally, the crowd may fail to do sufficiently well when its members are insufficiently independent from one another. There are convincing formal (Page, 2007) and empirical (Lorenz et al., 2011) reasons for thinking that the wisdom of the crowds effect requires that individuals' judgments are *to some degree* informationally independent, and that increased independence increases crowd accuracy.³² Intuitively, wisdom of the crowds phenomena work by pooling the private information people possess, which—when aggregated—is greater than the information possessed by any individual in the group. However, if individuals fail to contribute private information (perhaps because they suppress their own views and copy what other people have thought, or merely signal boost information they've come across elsewhere), this will tend to negate the effect.

³⁰ This thought is often comically captured by the popular "what people think is hard vs. what's actually hard" genre of video.

³¹ We are grateful to an anonymous reviewer for pointing out this concern.

³² See Dietrich & Spiekermann (2013) for discussion of how to weaken independence requirements.

We emphasize that complete epistemic independence is *not* required for wisdom of the crowds phenomena to arise. Individual evaluators can be epistemically correlated, drawing on some shared information, without breaking the wisdom of the crowds to identify experts in some contexts. In the extreme of complete overlapping information, no wisdom of the crowds can arise, and so it is important to ensure that there is some measure of independence, even if it does not need to be complete.

Moreover, there are some ways that social media algorithms reduce the independence of individual judges of content. The content that is most likely to be prioritized in search and recommendation will be content that has previously been consumed and judged favorably by others (or perhaps by bots or troll farms³³), and those judgments will often be obvious to subsequent consumers (e.g., visible view and reaction metrics; Mansoury et al., 2020). Thus, people will often lack full independence when consuming content, for they will often have some idea of whether the content is popular and whether people have positively or negatively reacted to it upon viewing.

We should expect reduced independence to have the following effects: people who become popular DIY content creators will tend to remain popular DIY content creators, and new content creators in that area will have a difficult time gaining popularity. Those who become popular because they offer good advice will be prioritized in search, and new people who consume their content—because it contains good advice—will tend to react positively to it and consume more content by those creators, resulting in continued prioritization in search. New content creators, on the other hand, will typically not have their content prioritized and may find it more difficult to generate significant views, thereby potentially causing content to stagnate and reducing the likelihood that new innovators in an area are able to gain popularity.³⁴ That is, epistemic correlations are likely to lead to increased entrenchment of those who were previously judged to be experts.

Importantly, however, this weakness is also a strength. In the early days of the internet, search engines and platforms often had a relatively uncontrolled feel. One never knew what content one might find, what kind of quality it would be, and so on. This had some advantages, occasionally resulting in refreshing, original, or delightful discoveries. But it also had significant costs when compared to the way platforms organize information today, as it was often very time consuming and tedious to find high quality DIY content (or relevant content of any kind). In contrast, while the sense of surprise and originality may, to some degree, be reduced on modern platforms, the ease with which one can find high quality relevant content plausibly compensates.

The lesson of this factor is that we should not expect the wisdom of the crowds as aggregated by recommendation algorithms to do particularly well at recognizing innovation. For that reason, even when the other factors may be sufficiently satisfied,

³³ For some of the ways that orchestrated campaigns and social bots can manipulate engagement systems see Ferrara (2016) and Shao et al., (2018), see also Stanley-Becker (2020 and Hao (2020) for examples and discussions of troll farms.

³⁴ For work on how this has affected online political communities see Hindman et al., (2003).

it may be best to rely on more traditional credentialing institutions if progress or innovation in a field is particularly important.

4 Explanations for Why People form Bad Beliefs

To sum up, we have argued that there are five epistemic factors about a subject matter that influence the likelihood that the crowd can be wise at expert identification on social media platforms. We have also argued that DIY content typically scores somewhat well on these factors, but not non-DIY content, especially (but not solely) content related to climate change, vaccines, and election safety. At this point, it is worth briefly stepping back and placing our account in the landscape of other accounts regarding why people form incorrect beliefs.

Our account explains when we should expect people to be sufficiently poor at evaluating online content that the crowd will be unable to reliably recognize experts. Notice that this is not yet an explanation for why people form bad beliefs. After all, it is possible that in many cases where people are likely to be poor at evaluating online content *on their own*, they nonetheless can recognize experts and defer to them. People may thereby be capable of forming accurate beliefs about the relevant propositions without the crowd thereby being able to serve as a reliable credentialing institution. Similarly, people may form beliefs for reasons that are not epistemic.

Nonetheless, it will be useful to consider how our account fits with more common explanations for bad beliefs. According to one such explanation, people's political identity plays an important role in explaining professions of bad belief. If responding to a proposition in a particular way is a marker of political identity, then many people may believe (or profess belief) in accordance with their political identity and not the truth (Enders et al., 2020), perhaps because of motivated reasoning (Kunda 1990; Lodge & Taber, 2013; Sinatra et al., 2014; Hart & Nisbet, 2012). We do not attempt to assess the merit of this view, but simply we note that—to the extent that it has merit—an additional factor should perhaps be added to our list.³⁵ In circumstances where people are likely to believe, profess belief, or consume content in accordance with political identity rather than truth (or in those circumstances where motivated reasoning is likely), we should expect the crowd to be less likely to be able to serve as a reliable credentialing institution.

Another explanation for bad belief appeals to people's social networks (Contessa, 2022; Levy, 2019; O'Connor & Weatherall, 2019). Much of what we know results from trusting our social networks, which can sometimes lead us astray. Our account is consistent with, and may even bolster, this style of explanation. We have identified a set of conditions under which people will be unable to reliably evaluate online content on their own. In these conditions, if they are to do well, they will have to rely on others. And, while we think recent work by education scholars suggests that it is possible for people to do this well (and that it's relatively easy to teach people how to do this well; Caulfield & Wineburg 2023), there is also good evidence that people (without these interventions) often don't do well at this (Breakstone et al.,

³⁵ Whether this is an *epistemic* factor, we leave to the side.

2021; Wineburg & McGrew, 2019). Thus, whenever a subject matter scores low on our factors, if people rely on their own judgments, then we should expect them to do poorly. Moreover, if someone relies on their social network for answers but that network is primarily composed of unreliable (and/or non-independent) individuals, then we should expect them to perform poorly.

Finally, a related explanation for bad beliefs appeals to the prevalence of deceptive content. This sort of explanation may have caused some readers to formulate the following worry about our view. Many people produce deceptive or misleading (or doubt-producing) content about vaccines, climate change, and election results (Oreskes & Conway, 2011; Hao 2020; Frenkel, 2021), but deceptive cake baking content is presumably much rarer. Perhaps this asymmetry in frequency of deceptive content, and not the epistemic factors we've proposed, explains why the wisdom of the crowds phenomenon works for DIY cases but not for medical content or climate science content.

While we agree that there is likely a deceptive content asymmetry that could play a role in explaining why people sometimes form problematic beliefs about non-DIY content, our account is in some sense prior to such impacts. We are interested in how people evaluate individual pieces of content, because it is these individual evaluations that add up to create the wisdom of the crowds phenomenon. That there is deceptive content does not yet tell us whether people will fall for it. Our account provides a partial explanation for *when* and *why* people will be bad at evaluating deceptive and non-deceptive content. And, when people are bad at evaluating content, we should expect them to be more likely to fall for deceptive content. Our account also helps explain why there is a deceptive content asymmetry in the first place. If deceptive content is unlikely to find much uptake because people are reliable (enough) judges of the quality of content in DIY cases and can therefore recognize when content is deceptive, then it will often be unprofitable to create such content.³⁶

5 Conclusion

We have argued for a novel wisdom of the crowd phenomenon. Specifically, we have argued that the crowd reliably picks out those with significant expertise in DIY content areas and can thus fulfill one important role of credentialing institutions. Much recent research on the epistemology of social media has had a decidedly negative tone. Social media platforms have been accused of shortening attention spans (Hari, 2022), contributing to the spread of misinformation (Myers, 2022) and the formation of filter bubbles (Sunstein, 2017), and increasing polarization (Barrett et al., 2021), and much more besides. In this flurry of work, it has perhaps been easy to overlook the epistemic achievements resulting from social media. We hope our work centers

³⁶ There are yet other explanations for bad beliefs. For example, some authors have argued that people form bad beliefs about human-caused climate change because they lack knowledge about what scientists believe. For an overview of some reasons to be skeptical of this position see Kovaka (2021). That paper also includes an interesting explanation for why a particular group of people, those who are pro-science, still sometimes deny humans' role in causing climate change. For reasons of space, we leave these explanations undiscussed.

one overlooked achievement. People are in a better position now than ever before to learn new hobbies, improve their expertise in a wide range of areas, and learn how to accomplish important and life-improving tasks because of social media.

Of course, the fact that this epistemic benefit arises in some cases does not thereby mean that social media platforms are well-designed or healthy (Amico-Korby et al., 2024). In particular, while social media platforms have made information accessible to an unprecedented degree, non-DIY information is often sorted by unreliable processes that prioritize and recommend popular and engaging content (Milli et al., 2023). Traditionally, social media platforms have dealt with this problem by ignoring the problem, or by using content moderation strategies, whether “hard” (e.g., removing content, deplatforming content creators) or “soft” (e.g., weighted filtering, sorting, recommending, and labeling of borderline problematic content and “authoritative” content). For example, platforms have occasionally made it impossible to search for certain pages, accounts, or types of content (Gillespie, 2018). While we take some issues with how platforms have enacted soft-content moderation in practice (and with the justifications they have given for it (Gillespie 2018)), we think this approach is the most promising.

We thus close by suggesting one practical soft-content moderation strategy for platforms to improve their epistemic environments: platforms should mark and prioritize (in recommendations and in search) non-DIY content that has been produced by experts who have been credentialed by public institutions.³⁷ This outsourcing approach would offer several benefits to platforms. First, they would not have to evaluate content themselves. Instead they, can offload the work of determining whether a person is an expert on a particular topic to public institutions that have been built to perform that work, and they can allow those experts to determine what content is appropriate to share. Second, they can build on (and borrow from) already-existing technical infrastructure for authentication (e.g., identity verification for famous individuals). Third, this approach need not require the removal of content produced by non-experts, thereby helping platforms avoid (some) difficult questions about when to take down content. Fourth, prioritizing content created by credentialed topic-experts will simplify finding high-quality information without infringing on people’s intellectual freedom to search out non-expert content if they so choose.³⁸ And so long as they rely on *public* credentials, the process of determining which accounts to prioritize will have some democratic oversight (e.g. via state-run licensing boards).

This proposal would make platforms epistemically more similar to the offline world. Offline, there are clearly marked places to find content from experts in a wide range of topics. For example, I can go to the doctor’s office for medical content, visit the mechanic for car repair advice, or go to a university to find climate scientists to

³⁷ For more in depth discussion of this strategy see (Amico-Korby et al., 2024). See also (Google 2023, Currin 2021) for more on how YouTube has implemented a closely related strategy for a narrow set of topics. And see Narayanan (2023) for a discussion of the effects of content demotion more generally.

³⁸ Importantly, this solution doesn’t fall victim to the pitfalls traditional approaches to these problems are plagued by Record & Miller (2022).

take a class from. Nonetheless, I have the freedom to instead consult a friend for medical advice, even if they lack medical credentials. Prioritizing content created by topic-experts for non-DIY content will make it such that the easiest to find information is of high quality, but without infringing on people's intellectual freedom to search out content that is not produced by experts if they so choose.³⁹

There are, of course, other strategies that platforms could pursue. For example, an educational platform might use research libraries' treatment of scientific and pseudoscientific content as a model. This strategy requires a more hands-on curation approach, and so platforms may be resistant to taking such an approach, although it may be more appealing if users are also given the option to turn off these curation features.⁴⁰ Ultimately, what's best for a platform will depend, to some degree on its epistemic function (Amico-Korby et al., 2024).⁴¹

One central worry platforms may have about these proposed changes is that it may risk reducing clicks or user engagement, and thus platforms' bottom lines. However, these strategies could be combined with engagement-based metrics; for instance, platforms could amplify high-engagement content from credentialed experts.⁴² However, we acknowledge that shifts to incorporate epistemic desiderata, including efforts to improve epistemic assessments, could raise monetary concerns for platforms. However, we think that the urgency of the problem provides reasons for platforms to take steps to improve how they sort non-DIY content despite monetary risks.⁴³

Funding Office of Naval Research.

³⁹ There are certain practical difficulties with distinguishing between DIY and non-DIY content. We address these in greater detail in Amico-Korby et al., (2024). This strategy also still permits bad actors (and sincere but misinformed actors) to continue to share and spread bad content on platforms. However, because high quality content will be prioritized in search and recommendations, the spread of problematic content should be reduced (Narayanan 2023).

⁴⁰ For a discussion of how some platforms have used a similar approach to handle pornography, see (Gillespie 2018, chapter 7). For discussions of how platforms might, and why there might be good reasons for platforms to, provide more user control over algorithms see Fukuyama et al. (2020), and the following amicus brief submitted to the supreme court during the *NetChoice* cases (Fukuyama et al., 2024).

⁴¹ Another possible companion strategy for platforms is outlined by Record & Miller (2022). They argue that platforms ought to nudge people to develop and adhere to certain recognized epistemic norms for posting.

⁴² Still, we should be cautious here. Avram et al., (2020) have raised concerns that engagement metrics increase vulnerability to misinformation (display of engagement metrics influences interaction with low-credibility content) and McLoughin et al., (2021) have demonstrated how outrage inducing misinformation increased engagement. While only using engagement metrics to sort content that meets high epistemic standards should help alleviate some of these concerns, platforms should approach these problems with slow-roll outs, careful testing, and openness to changing course. One particular concern is that experts may attempt to tailor their content in such a way that it gets amplified, at the cost of epistemic quality.

⁴³ For more on platforms obligations to create epistemically healthy environments, see Marin (2022) and Amico-Korby et al. (2024).

Declarations

Conflict of interest All authors contributed to all aspects of this paper. This research was partially supported by a grant to DD from the Office of Naval Research. The authors have no financial or proprietary interests in any material discussed in this article. We are grateful to Jennifer Rose Carr, and audiences at UC San Diego and Cal Poly SLO for their many helpful comments.

References

- “A Few Factors in Search Ranking—YouTube Help.” Accessed April 9, 2024. https://web.archive.org/web/20150329045851/https://support.google.com/youtube/answer/6046978?hl=en&ref_topic=6046759.
- “Get Info on Health-Related Content - YouTube Help.” Accessed March 11, 2024. <https://support.google.com/youtube/answer/9795167?hl=en>.
- American Library Association. “Labeling and Rating Systems Q&A | Advocacy, Legislation & Issues,” June 2024. <https://www.ala.org/advocacy/intfreedom/labelingratingqa>.
- Wikipedia: Reliable Sources/Perennial Sources. In Wikipedia, April 1, 2024. https://en.wikipedia.org/w/index.php?title=Wikipedia:Reliable_sources/Perennial_sources&oldid=1216627486#Sources.
- YouTube Health—Accessible Public Health Information. Accessed April 2, 2024. <https://health.youtube.com>.
- Alba, D. “The Latest Covid Misinformation Star Says He Invented the Vaccines (Published 2022),” April 3, 2022. <https://www.nytimes.com/2022/04/03/technology/robert-malone-covid.html>.
- Alfano, M., Fard, A. E., Adam Carter, J., Clutton, P., & Klein, C. (2021). Technologically scaffolded atypical cognition: the case of Youtube’s recommender system. *Synthese*, 199, 835–858. <https://doi.org/10.1007/s11229-020-02724-x>
- Amico-Korby, D., Harrell, M. & Danks, D. (2024). Environmental epistemology. *Synthese*, 203, 81. <https://doi.org/10.1007/s11229-024-04504-3>
- Amico-Korby, D., Harrell, M., & Danks, D. (2024). *Building Epistemically Healthier Platforms*. Episteme.
- Anderson, E. (2011). Democracy, public policy, and lay assessments of scientific testimony. *Episteme*, 8(2), 144–164. <https://doi.org/10.3366/epi.2011.0013>
- Askin, A., Sengul, L., & Tosun, A. (2020). YouTube as a source of information for transcranial magnetic stimulation in stroke: a quality, reliability and accuracy analysis. *Journal of Stroke and Cerebrovascular Diseases: THE Official Journal of National Stroke Association*, 29(12), 105309. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105309>
- Avram, M., Nicholas M., Sameer P., & Filippo M. “Exposure to Social Engagement Metrics Increases Vulnerability to Misinformation | HKS Misinformation Review,” 2020. <https://misinformreview.hks.harvard.edu/article/exposure-to-social-engagement-metrics-increases-vulnerability-to-misinformation/>.
- Ayo-Ajibola, O., Davis, R. J., Theriault, C., Lamb, C., Choe, D., Lin, M. E., Angell, T. E., & Kwon, D. I. (2024). Evaluation of YouTube as a source for Grave’s disease information: Is high-quality guideline-based information available? *OTO Open*, 8(1), e118. <https://doi.org/10.1002/oto2.118>
- Barrett, P., Justin H., & Grant Sims J. “Fueling the Fire: How Social Media Intensifies U.S. Political Polarization--And What Can Be Done About It.” NYU Stern Center for Business and Human Rights, 2021. <https://www.stern.nyu.edu/experience-stern/faculty-research/fueling-fire-how-social-media-intensifies-u-s-political-polarization-and-what-can-be-done-about-it>.
- Barrett, P. M., & Justin H. “A Platform ‘Weaponized’: How YouTube Spreads Harmful Content--And What Can Be Done About It.” NYU Stern Center for Business and Human Rights, June 2022. https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/62a38fc022745a7274601da0/1654886337000/NYU+CBHR+YouTube_Final_June10.pdf.
- Barrett, T. “Inhofe Brings Snowball on Senate Floor as Evidence Globe Is Not Warming | CNN Politics.” CNN Politics 2015. Accessed March 11, 2024. <https://www.cnn.com/2015/02/26/politics/james-inhofe-snowball-climate-change/index.html>.
- Bartlett, T. “The Vaccine Scientist Spreading Vaccine Misinformation.” The Atlantic, August 12, 2021. <https://www.theatlantic.com/science/archive/2021/08/robert-malone-vaccine-inventor-vaccine-skeptic/619734/>.

- Beauchamp, Z. “Alex Jones, Pizzagate Booster and America’s Most Famous Conspiracy Theorist, Explained - Vox,” 2016. <https://www.vox.com/policy-and-politics/2016/10/28/13424848/alex-jones-infowars-prisonplanet>.
- Bella, T. “A Vaccine Scientist’s Discredited Claims Have Bolstered a Movement of Misinformation.” Washington Post, January 24, 2022. <https://www.washingtonpost.com/health/2022/01/24/robert-malone-vaccine-misinformation-rogan-mandates/>.
- Benkler, Y., Tilton, C., Etlng, B., Roberts, H., Clark, J., Faris, R., Kaiser, J., & Schmitt, C. (2020). Mail-in voter fraud: anatomy of a disinformation campaign. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3703701>
- Birch, E. M., Leziak, K., Jackson, J., Dahl, E., Niznik, C. M., & Yee, L. M. (2022). Content quality of YouTube videos about gestational diabetes: systematic evaluation. *JMIR Diabetes*, 7(2), e30156. <https://doi.org/10.2196/30156>
- blog.youtube. “The Four Rs of Responsibility, Part 2: Raising Authoritative Content and Reducing Borderline Content and Harmful Misinformation.” Accessed April 1, 2024. <https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-raise-and-reduce/>.
- Bradford, G. (2017). Problems for perfectionism. *Utilitas*, 29(3), 344–364. <https://doi.org/10.1017/S0953820816000418>
- Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Student’s civic online reasoning: a national portrait. *Educational Researcher*, 50(8), 505–515. <https://doi.org/10.3102/0013189X211017495>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Brink, D. O. (2013). *Mill’s Progressive Principles*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672141.001.0001>
- Brink, D. O. (2019). Normative perfectionism and the kantian tradition. *Philosophers’ Imprint*, 19(45).
- Brown, M., Bisbee, J., Lai, A., Bonneau, R., Nagler, J., & Tucker, J. A. (2022). Echo chambers, rabbit holes, and algorithmic bias: how YouTube recommends content to real users. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4114905>
- Bump, P. “Analysis | ‘2000 Mules,’ a Key Piece of Election Misinformation, Has Its Day in Court.” Washington Post, October 3, 2023. <https://www.washingtonpost.com/politics/2023/10/03/judge-rules-defamation-lawsuit-against-2000-mules-can-proceed/>.
- Campbell, D. R. (2024). Not just a tool: why social-media use is bad and bad for us, and the duty to quit. *Journal of Global Ethics*. <https://doi.org/10.1080/17449626.2024.2339451>
- Caulfield, M., & Wineburg, S. (2023). *Verified: How to Thing Straight, Get Duped Less, and Make Better Decisions about What to Believe Online*. Chicago: The University of Chicago Press.
- Celik, H., Polat, O., Ozcan, C., Camur, S., Kilinc, B. E., & Uzun, M. (2020). Assessment of the quality and reliability of the information on rotator cuff repair on YouTube. *Orthopaedics and Traumatology, Surgery and Research: OTSR*, 106(1), 31–34. <https://doi.org/10.1016/j.otsr.2019.10.004>
- Cercone, J. “The Real COVID-19 Death Count in England Not Overstated.” politifact. Accessed March 11, 2024. <https://www.politifact.com/factchecks/2022/jan/24/youtube-videos/no-death-totals-covid-19-england-have-not-been-ove/>.
- Coaston, J. Alex Jones Banned from YouTube, Facebook, and Apple, Explained—Vox. Vox, August 6, 2018. <https://www.vox.com/2018/8/6/17655658/alex-jones-facebook-youtube-conspiracy-theories>.
- Contessa, G. (2022). It takes a village to trust science: towards a (thoroughly) social approach to public trust in science. *Erkenntnis*, 88(7), 2941–2966. <https://doi.org/10.1007/s10670-021-00485-8>
- Covington, P., Jay A., & Emre S. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–98. Boston Massachusetts USA: ACM, 2016. <https://doi.org/10.1145/2959100.2959190>.
- Crutchfield, C. R., Frank, J. S., Anderson, M. J., Trofa, D. P., & Sean Lynch, T. (2021). A systematic assessment of YouTube content on femoroacetabular impingement: an updated review. *Orthopaedic Journal of Sports Medicine*, 9(6), <https://doi.org/10.1177/23259671211016340>
- Currin, G. “YouTube’s Plan to Showcase Credible Health Information Is Flawed, Experts Warn.” Scientific American, 2021. <https://www.scientificamerican.com/article/youtubes-plan-to-showcase-credible-health-information-is-flawed-experts-warn/>.
- Dautrich, K. “The Future of the First Amendment: 2018 Survey of High School Students and Teachers.” Knight Foundation. Accessed October 8, 2024. <https://knightfoundation.org/reports/the-future-of-the-first-amendment-2018/>.

- de Zúñiga, G., Homero, B. W., & Ardèvol-Abreu, A. (2017). Effects of the news-finds-me perception in communication: social media use implications for news seeking and learning about politics. *Journal of Computer-Mediated Communication*, 22(3), 105–123. <https://doi.org/10.1111/jcc4.12185>
- Dietrich, F., & Spiekermann, K. (2013). Epistemic democracy with defensible premises. *Economics & Philosophy*, 29(1), 87–120. <https://doi.org/10.1017/S0266267113000096>
- DiResta, R. “The Web’s Recommendation Engines Are Broken. Can We Fix Them? | WIRED.” Wired.com, April 11, 2018. <https://www.wired.com/story/creating-ethical-recommendation-engines/>.
- Enders, A. M., Uscinski, J. E., Klofstad, C., & Stoler, J. (2020). The different forms of COVID-19 misinformation and their consequences. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-48>
- Etzel, C. M., Bokshan, S. L., Forster, T. A., & Owens, B. D. (2022). A quality assessment of YouTube content on shoulder instability. *The Physician and Sportsmedicine*, 50(4), 289–294. <https://doi.org/10.1080/00913847.2021.1942286>
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>
- Knight Foundation. “The Future of the First Amendment: 2018 Survey of High School Students and Teachers.” Accessed October 8, 2024. <https://knightfoundation.org/reports/the-future-of-the-first-amendment-2018/>.
- Frenkel, S. “The Most Influential Spreader of Coronavirus Misinformation Online (Published 2021),” July 24, 2021. <https://www.nytimes.com/2021/07/24/technology/joseph-mercola-coronavirus-misinformation-online.html>.
- Fukuyama, F., Barak R., Ashish G., Marietje S., Roberta R. K., & Douglas M. “Report of the Working Group on Platform Scale,” November 17, 2020. <https://cyber.fsi.stanford.edu/publication/report-working-group-platform-scale>.
- Fukuyama, F., Daphne K., & Jack M. B. Brief of Amicus Curiae Francis Fukuyama In Support of Respondents in No. 22-277 and Petitioners in No. 22-55 Ashley Moody, Attorney General of Florida, ET AL., Petitioners, v. NetChoice LLC, DBA NetChoice, et al., Respondents and NetChoice LLC, DBA NetChoice, et al., Petitioners v. Ken Paxton, Attorney General of Texas, Respondent., No. 22-277, 22-555. Accessed March 11, 2024.
- Funk, C., Alec T., Brian K., & Giancarlo P. “Americans’ Largely Positive Views of Childhood Vaccines Hold Steady.” *Pew Research Center Science & Society* (blog), May 16, 2023. <https://www.pewresearch.org/science/2023/05/16/americans-largely-positive-views-of-childhood-vaccines-hold-steady/>.
- Gillespie, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven [Connecticut]: Yale University Press, 2018.
- Gillespie, T. (2022). Do not recommend? Reduction as a form of content moderation. *Social Media + Society*. <https://doi.org/10.1177/20563051221117552>
- Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1), 85–110. <https://doi.org/10.2307/3071090>
- Goodrow, C. “On YouTube’s Recommendation System.” blog.youtube. Accessed April 2, 2024. <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>.
- Gorski, D. “Conspiracy Theories about Monkeypox: Déjà vu All over Again or Same as It Ever Was? | Science-Based Medicine,” August 8, 2022. <https://sciencebasedmedicine.org/conspiracy-theories-about-monkeypox/>.
- Guerrero, A. A. (2016). Living with ignorance in a world of experts. In R. P. (Ed.), *Perspectives on Ignorance from Moral and Social Philosophy*. New York: Routledge.
- Hao, K. “Troll Farms Reached 140 Million Americans a Month on Facebook before 2020 Election | MIT Technology Review.” MIT Technology Review, September 16, 2021. <https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election/>.
- Hari, J. “Your Attention Didn’t Collapse. It Was Stolen.” *the Guardian*, January 2, 2022. <https://www.theguardian.com/science/2022/jan/02/attention-span-focus-screens-apps-smart-phones-social-media>.
- Amico-Korby, Harrell, M.D., Danks, D. “Using Verification to Help Social Media Users Recognize Experts | TechPolicy.Press.” Tech Policy Press, July 17, 2023. <https://techpolicy.press/using-verification-to-help-social-media-users-recognize-experts>.
- Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: how motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, 39(6), 701–723. <https://doi.org/10.1177/0093650211416646>

- Haugan, T. “Flat Earth, Bigotry, and Bias: Should the Library Be Neutral? | UC Berkeley Library.” Accessed April 2, 2024. <https://www.lib.berkeley.edu/about/news/neutral>.
- Hindman, M, Kostas T, & Judy A. J. “Googlearchy’: How a Few Heavily-Linked Sites Dominate Politics on the Web. *Annual Meeting of the American Political Science Association, Philadelphia, PA.*, 2003.
- Jaworska, S., Goodman, M. K., & Gibas, I. (2024). The making of #CovidTwitter: Who were the loudest “Covid Influencers” and what did they say about the COVID-19 pandemic? *Social Media + Society*. <https://doi.org/10.1177/20563051231222240>
- Kovaka, K. (2021). Climate change denial and beliefs about science. *Synthese*, 198(3), 2355–2374. <https://doi.org/10.1007/s11229-019-02210-z>
- Kreps, S, & Doug K. “KK_Covid_Misinformation_OSF_10 June*.Pdf.” *OSF*, June 10, 2020. <https://osf.io/https://osf.io/ceagm>.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Latkin, C., Dayton, L., Miller, J., Eschliman, E., Yang, J., Jamison, A., & Kong, X. (2023). Trusted information sources in the early months of the COVID-19 pandemic predict vaccination uptake over one year later. *Vaccine*, 41(2), 573–580. <https://doi.org/10.1016/j.vaccine.2022.11.076>
- Leiter, B. (2021). The epistemology of the internet and the regulation of speech in America. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3939948>
- Levy, N. (2019). Due deference to denialism: explaining ordinary people’s rejection of established scientific findings. *Synthese*, 196(1), 313–327. <https://doi.org/10.1007/s11229-017-1477-x>
- Lewis, P. “‘Fiction Is Outperforming Reality’: How YouTube’s Algorithm Distorts Truth.” *The Guardian*, February 2, 2018, sec. Technology. <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>.
- Lodge, M., & Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139032490>
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348. <https://doi.org/10.1038/s41562-021-01056-1>
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025. <https://doi.org/10.1073/pnas.1008636108>
- Mamo, A., Szeto, M. D., Mirhossaini, R., Fortugno, A., & Dellavalle, R. P. (2021). Tetrahydrocannabinol and skin cancer: analysis of YouTube videos. *JMIR Dermatology*, 4(1), e26564. <https://doi.org/10.2196/26564>
- Mansoury, M, Himan A, Mykola P, Bamshad M, & Robin B. “Feedback Loop and Bias Amplification in Recommender Systems.” arXiv, July 25, 2020. <https://doi.org/10.48550/arXiv.2007.13019>.
- Marin, L. (2022). How to do things with information online. A conceptual framework for evaluating social networking platforms as epistemic environments. *Philosophy and Technology*, 35(77).
- Masterton, G., & Olsson, E. J. (2018). From impact to importance: the current state of the wisdom-of-crowds justification of link-based ranking algorithms. *Philosophy & Technology*, 31(4), 593–609. <https://doi.org/10.1007/s13347-017-0274-2>
- McGrady, R. “What We Discovered on ‘Deep YouTube.’” *The Atlantic* (blog), January 26, 2024. <https://www.theatlantic.com/technology/archive/2024/01/how-many-videos-youtube-research/677250/>.
- McGrady, R., Zheng, K., Curran, R., Baumgartner, J., & Zuckerman, E. (2023). Dialing for videos: a random sample of YouTube. *Journal of Quantitative Description: Digital Media*. <https://doi.org/10.51685/jqd.2023.022>
- McKenna, R. *Non-Ideal Epistemology—Robin McKenna—Oxford University Press*. Oxford University Press, 2023. <https://global.oup.com/academic/product/non-ideal-epistemology-9780192888822?cc=us&lang=en&>.
- McLoughlin, K. L., Brady, W. J., & Crockett, M. J. (2021). The role of moral outrage in the spread of misinformation. *PubPub*. <https://doi.org/10.1037/tms0000136>
- Meyerson, E. “YouTube Now: Why We Focus on Watch Time.” blog.youtube, August 10, 2012. <https://blog.youtube/news-and-events/youtube-now-why-we-focus-on-watch-time/>.
- Millgram, E. *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*. Oxford University Press, 2015.

- Milli, S., Micah C., Yike W., Sashrika P., Sebastian Z., & Anca D. D. "Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media." arXiv.org, May 26, 2023. <https://arxiv.org/abs/2305.16941v5>.
- United States: Justice Department, Robert S. Mueller, United States: District Court (District of Columbia), Robert S. Mueller, Internet Research Agency, LLC, United States, and United States: Department of Justice.: Special Counsel's Office., "United States of America v. Internet Research Agency Llc [and 15 Others], Defendants: Case 1:18-Cr-00032-Dlf." Justice Department, January 1, 2018. <https://purl.fdlp.gov/GPO/gpo89499>.
- Murugiah, K., Vallakati, A., Rajput, K., Sood, A., & Challa, N. R. (2011). YouTube as a source of information on cardiopulmonary resuscitation. *Resuscitation*, 82(3), 332–334. <https://doi.org/10.1016/j.resuscitation.2010.11.015>
- Myers, S. L. "Twitter and TikTok Lead in Amplifying Misinformation, Report Finds - The New York Times," October 13, 2022. <https://www.nytimes.com/2022/10/13/technology/misinformation-integrity-institute-report.html>.
- Narayanan, A. "Understanding Social Media Recommendation Algorithms." Knight First Amendment Institute at Columbia University, Mrrach 2023. <http://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms>.
- Newton, C. "YouTube Deletes Alex Jones' Channel for Violating Its Community Guidelines - The Verge," August 6, 2018. <https://www.theverge.com/2018/8/6/17656708/youtube-alex-jones-infowars-account-deleted-facebook-apple-spotify>.
- Nielsen, J. "How Long Do Users Stay on Web Pages?" Nielsen Norman Group, September 11, 2011. <https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/>.
- Nguyen, C. T. (2020a). Cognitive islands and runaway echo chambers: Problems for epistemic dependence on experts. *Synthese*, 197(7), 2803–2821. <https://doi.org/10.1007/s11229-018-1692-0>
- Nguyen, C. T. (2020b). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- O'Connor, C., & James O. W. *The Misinformation Age: How False Beliefs Spread*. New Haven London: Yale University Press, 2019.
- Oreskes, N., & Conway, E. M. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Climate Change* (Paperback). Nachdruck. New York London Oxford New Delhi Sydney: Bloomsbury.
- Özdemir, E. E., Borman, P., Civelek, G. M., & Umaroğlu, M. M. (2023). YouTube as a source of information on lipedema: property, quality, and reliability assessment. *Lymphatic Research and Biology*, 21(4), 403–409. <https://doi.org/10.1089/lrb.2022.0028>
- Page, S. E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. 3. print., and 1. paperback print., with A new preface. Princeton, NJ: Princeton Univ. Press, 2007.
- Park, C. S., & Kaye, B. K. (2021). What's This? Incidental Exposure to News on Social Media, News-Finds-Me Perception, News Efficacy, and News Consumption. In F. Shen (Ed.), *Social Media News and Its Impact* (pp. 98–121). New York: Routledge. <https://doi.org/10.4324/9781003179580-6>
- Pasquini, G., Alison S., Alec T., & Cary F. Why Some Americans Do Not See Urgency on Climate Change. *Pew Research Center Science & Society* (blog), August 9, 2023. <https://www.pewresearch.org/science/2023/08/09/why-some-americans-do-not-see-urgency-on-climate-change/>.
- Pew Research Center. "4. Voters' Reflections on the 2020 Election." *Pew Research Center—U.S. Politics & Policy* (blog), January 15, 2021. <https://www.pewresearch.org/politics/2021/01/15/voters-reflections-on-the-2020-election/>.
- Pew Research Center. "Trust in Media Archives," October 27, 2022. <https://www.pewresearch.org/topic/news-habits-media/media-society/media-attitudes/trust-in-media/>.
- Pew Research Center. "Social Media and News Fact Sheet," September 17, 2024. <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>.
- Porter, E., Velez, Y., & Wood, T. J. (2023). Correcting COVID-19 vaccine misinformation in 10 countries. *Royal Society Open Science*. <https://doi.org/10.1098/rsos.221097>
- Record, I., & Miller, B. (2022a). People, posts, and platforms: reducing the spread of online toxicity by contextualizing content and setting norms. *Asian Journal of Philosophy*. <https://doi.org/10.1007/s44204-022-00042-2>
- Record, I., & Miller, B. (2022b). Wrong on the Internet: Why Some Common Prescriptions for Addressing the Spread of Misinformation Online Don't Work. *Communiqué*, 105, 22–27.

- Rep. Malinowski, Tom [D-NJ-7. “Text—H.R.2154—117th Congress (2021–2022): Protecting Americans from Dangerous Algorithms Act.” Legislation, March 24, 2021. 2021-03-23. <https://www.congress.gov/bill/117th-congress/house-bill/2154/text>.
- Reuters Fact Check. Does ‘2000 Mules’ Provide Evidence of Voter Fraud in the 2020 U.S. Presidential Election? Reuters, May 27, 2022. <https://www.reuters.com/article/fidUSL2N2XJ0OQ/>.
- Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to ‘ranking cultures’: investigating the modulation of visibility in YouTube search results. *Convergence: the International Journal of Research into New Media Technologies*, 24(1), 50–68. <https://doi.org/10.1177/1354856517736982>
- Rini, R. (2021). Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology. In E. Edenberg & M. Hannon (Eds.), *Political Epistemology* (pp. 31–48). Oxford University Press/Oxford. <https://doi.org/10.1093/oso/9780192893338.003.0003>
- Roose, K, Mike I, & Sheera F. Facebook Struggles to Balance Civility and Growth. *The New York Times*, November 24, 2020, sec. Technology. <https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html>.
- Schuster, N., & Lazar, S. (2024). Attention, moral skill, and algorithmic recommendation. *Philosophical Studies*. <https://doi.org/10.1007/s11098-023-02083-6>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*. <https://doi.org/10.1038/s41467-018-06930-7>
- Sinatra, G. M., Kienhues, D., & Hofer, B. K. (2014). Addressing challenges to public understanding of science: epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist*, 49(2), 123–138. <https://doi.org/10.1080/00461520.2014.916216>
- Solomon, M. (2006). Groupthink versus the wisdom of crowds: the social epistemology of deliberation and dissent. *Southern Journal of Philosophy*, 44(S1), 28–42. <https://doi.org/10.1111/j.2041-6962.2006.tb00028.x>
- Stanley-Becker, I. Pro-Trump Youth Group Enlists Teens in Secretive Campaign Likened to a ‘Troll Farm,’ Prompting Rebuke by Facebook and Twitter. Washington Post, September 15, 2020. https://www.washingtonpost.com/politics/turning-point-teens-disinformation-trump/2020/09/15/c84091ae-f20a-11ea-b796-2dd09962649c_story.html.
- Sullivan, E., Max S., Ignaz R., Wouter M., Scott C., Bettina S., & Mark A. (2020). Can Real Social Epistemic Networks Deliver the Wisdom of Crowds? In L. Tania, K. Joshua, & N. Shaun (Eds.), *Oxford Studies in Experimental Philosophy*. Oxford University Press, Oxford. <https://doi.org/10.1093/oso/9780198852407.003.0003>
- Sunstein, C. R. (2017). *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- Surowiecki, James. *The Wisdom of Crowds*. New York: Anchor Books, a division of Random House, 2005. <https://doi.org/10.5555/1095645>.
- Tang, K., Azhar, U., Babar, M., Ahmed, A., Aaron, Oh., Day, W., Harb, H., & Chan, F. J. (2022). Assessing the quality of youtube videos on adhesive capsulitis. *Cureus*, 14(7), e27406. <https://doi.org/10.7759/cureus.27406>
- Tang, L., Fujimoto, K., Amith, M., Cunningham, R., Costantini, R. A., York, F., Xiong, G., Boom, J. A., & Tao, C. (2021). Down the rabbit hole’ of vaccine misinformation on YouTube: network exposure study. *Journal of Medical Internet Research*, 23(1), e23262. <https://doi.org/10.2196/23262>
- Tenbarge, K. “New York Passes Legislation That Would Ban ‘addictive’ Social Media Algorithms for Kids.” NBC News, June 7, 2024. <https://www.nbcnews.com/tech/social-media/new-york-passes-legislation-ban-addictive-social-media-algorithms-kids-rcna155470>.
- Thorburn, L. “How Platform Recommenders Work.” *Understanding Recommenders* (blog), November 23, 2022. <https://medium.com/understanding-recommenders/how-platform-recommenders-work-15e260d9a15a>.
- Thorn, P. D., & Schurz, G. (2012). Meta-induction and the wisdom of crowds. *Analyse & Kritik*, 34, 339–365.
- Tufekci, Z. “Opinion|YouTube, the Great Radicalizer.” *The New York Times*, March 10, 2018, sec. Opinion. <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>
- United States: Justice Department, Robert S. Mueller, United States: District Court (District of Columbia), Robert S. Mueller, Internet Research Agency, LLC, United States, and United States: Department of Justice.: Special Counsel’s Office., “United States of America vs. Internet Research

- Agency Llc [and 15 Others], Defendants: Case 1:18-Cr-00032-Dlf.” Justice Department, January 1, 2018. <https://purl.fdlp.gov/GPO/gpo89499>
- Van Swol, L., Lee, S., & Hutchins, R. (2022). The banality of extremism: the role of group dynamics and communication of norms in polarization on January 6. *Group Dynamics: Theory, Research, and Practice*, 26(3), 239–251. <https://doi.org/10.1037/gdn0000180>
- Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: reading less and learning more when evaluating digital information. *Teachers College Record: the Voice of Scholarship in Education*, 121(11), 1–40. <https://doi.org/10.1177/016146811912101102>
- Yi, S. K., Michael, M. S., Lee, M. D., & Dry, M. J. (2012). The wisdom of the crowd in combinatorial problems. *Cognitive Science*, 36(3), 452–470. <https://doi.org/10.1111/j.1551-6709.2011.01223.x>
- Zannettou, S. (2021). “I won the election!”: an empirical analysis of soft moderation interventions on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 15, 865–876. <https://doi.org/10.1609/icwsm.v15i1.18110>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.