INTRODUCTION

In his notable work *An Enquiry into Human Understanding*, influential eighteenth-century Scottish philosopher David Hume wrote the following, “a wise man proportions his belief to the evidence.”¹ As rational creatures existing in a civil society, we are continually hunting for new information in order to process and update our beliefs regarding the world within which we live. This is an essential human adaptation that has played a significant role in our survival and progress as a species. However, this process of encountering and verifying knowledge about the world around us, what philosophers study as epistemology, does not occur in a vacuum. Our ability to successfully evaluate evidence and form what epistemologists describe as “true belief” has as much to do with our social connections as it does with our individual psychology.²

As we have progressed as a civilization to vastly complex and interwoven informational and communication networks resultant from the implementation and expansion of the internet, mobile and wireless technology, and the explosion of social media, the progressive development of a number of new epistemological obstacles have appeared. This newly treacherous cognitive path towards true belief only worsens when one considers that a significant portion of our information seeking behavior is now completely under the direction of automated algorithmic systems. These systems have garnered extensive epistemic trust when the intent of the majority of their implementations falls outside of the altruistic scope of providing verifiably true information to serve the public good. This paper aims to discuss our typical misperception of what algorithms are, in comparison to what they are in reality, and how algorithmic systems intersect with our information seeking behavior, and our epistemological verification of knowledge, and how this interaction may contribute to widespread epistemic confusion and crisis.

ALGORITHMS

Algorithms are typically defined by computer scientists as “a description of the method by which a task is to be accomplished.”³ One can often hear this echoed in the exemplar of a typical cake baking recipe being described as an algorithm. This is an easily digestible description that is also almost completely removed from the dense algorithmic systems we now interact with on a near constant basis. We typically imagine software as a single entity and tend to be less cognizant of the processes occurring underneath the polished user interface we are presented. One singular algorithm is a minute component of the whole of a given software system. The algorithm itself is “independent of programming languages, and independent of the machine that executes the program containing the

---

¹Ian O’Hara is Assistant Professor and Research & Instruction Librarian, Weinberg Memorial Library, University of Scranton, ian.ohara@scranton.edu
Software is quite literally, the gathering and assemblage of multiple code files into an “executable” file, it is fundamentally a layered entity that is dependent on a myriad of interrelations and devices in order to function as intended. This inherent complexity engenders a fundamental misunderstanding related to what an algorithm is, how an algorithm functions, and how that algorithm interacts with the remaining assemblage of components within a software product to transmute given input data and return some desired output to the user.

The ubiquity of algorithmic systems in our daily lives is clearly observable. Information technology has been so vastly implemented and adopted that we can view the process as a colonization of not only our attention, which is most superficially apparent, but our intellectual processes, and nearly every other aspect of how we move through and function within the world. The widespread adoption of technological tools has occurred in a shockingly rapid fashion in which there may be gaps in the critical understanding of the potential implications of this complete acquiescence to a singular reliance on software systems. Our transference of information management has occurred critical examination and intellectual grasp of inherent functionality.

In order for algorithmic systems to function all data inputs must be reduced to mathematical proxies. In other words, complex real-world phenomena must be simplified, reduced, and abstracted at some point into a form that can be understood and acted upon by a machine. There must be a choice in the design of a given algorithm within a larger algorithmically driven software system regarding how to model these complex real-world phenomena mathematically in order to translate for the machine. For our purposes here, a mathematical model, computationally speaking can be understood as a metaphor, an explanatory simplification that usefully distills the complexity of reality but may also distort it.

The implications of this effect are quite evident when we are thinking about algorithmically driven information systems. In order to understand how the multiplicity of algorithms within a given information system translate, and potentially manipulate and distort our intended search, we as users would need some insight into the implementation details of a system, and subsequently some reasonable level of knowledge regarding the functionality of that information system. Unfortunately, most of this information is proprietary, rendering the majority of commercially available information systems to be often described as black boxes, in that they produce material effects in the world by providing the user with some set of information, while revealing very little about how that information has been located and provided to the user. The user of a given search system, for example, Google search, is faced with an extremely simple user interface that masks and abstracts away an undecipherable level of complexity and algorithmic decision making. This notion of the inability to intellectually grasp the functionality of an algorithmically driven information system is further complicated by the fact that the implementation of software systems is not static, but constantly evolving in order to “improve” system performance. This renders the potential for reverse engineering in order to explore system functionality nearly impossible as the implementation details of proprietary systems is a perpetual moving target.

Algorithmic systems and their relationship to information consumption have extended beyond just information seeking to encompass the utilization of software tools across nearly every facet of our society. Algorithmic logic has become pervasive in fields as finance, transportation, higher education, journalism, security, surveillance, popular culture, and in the media industry. Nowhere is this more clearly exemplified than the Facebook news feed. Here, algorithms exert ranking, and ordering processes upon Facebook posts within a given network of individual profiles, these algorithms are built to make decisions about inclusion, exclusion and ordering based on criteria that is intended to increase user engagement. Think for a moment about the diversity of subject areas of information that are found on a Facebook news feed. Everything exists here from political memes, to shares of articles from reputable scholarly journals, to typically relied upon news sources, to opinion pieces from verifiably agenda-based internet media sites, to demonstrable propagandizing. How many Facebook users can pinpoint with any measure of specificity how and why a given post has arrived in their news feed?

This inherent complexity within algorithmically driven systems, coupled with the overt obfuscation of implementation details contributes to a widespread tension in what we perceive these systems to be, what we need them to be, and what they actually are. Software systems have traditionally been put forth and considered to be computational and therefore objective and neutral, but they are making constant value laden decisions that are directly descendant of the values of individual software developers, corporate entities, political actors, and problematic socio-cultural hierarchies. Algorithmic systems tend to be inherently trusted, while the intellec-
tual control over how they have located, ranked, sorted, and displayed a given information set to us is woefully misunderstood.16

**EPISTEMOLOGY & EPISTEMIC CRISIS**

Philosophers describe the concept of epistemology as the study of the nature of knowledge, and justification: in particular, the study of the defining components, the substantive conditions or sources, and the limits of knowledge and justification.17 Plato, Kant, and other philosophical luminaries hold an influential view that propositional knowledge has three individually necessary and jointly sufficient components for acceptance: justification, truth, and belief.18 In this traditional philosophical understanding, knowledge is not simply limited to belief, but in order for a belief of any kind to become what we understand to be knowledge, we must have adequate indication that a known proposition is in fact true.19 We can think of this adequate indication of truth as evidence for the justification of said belief, and it follows that if there is an adequate amount of evidence for a held belief to satisfy the justification of such a belief, we arrive at true knowledge.20 Put simply, our heuristics related to individual epistemology are concerned with the source of information we are evaluating, how and why that information was produced, and the authority of the producers of the item of information being evaluated for justification.

Social epistemology extends this individual epistemological heuristic a bit further beyond the knowledge gathering process of the individual. Philosophers interested in social epistemology are primarily concerned with the evidence gathering processes of groups of individuals all participating in what can be described as an “epistemic system”.21 Clearly, this has significant implications related to our modern, intensely interconnected world. We are concerned here with epistemological consequentialism, or in other words, the evaluation of the impact of social systems broadly on epistemic outcomes.22 Increasingly epistemologists have become interested in the relationship between social influences and proper epistemic behavior.23 As a society, our epistemology has become significantly reliant on the concept of sharing information. In parallel to this interconnectedness, we have also begun to rely almost exclusively on algorithmic systems for the creation of a platform to facilitate a significant amount of our social lives, our interactions socially have begun to take place more frequently in online spaces that are managed and curated algorithmically.

From our high-level discussion of individual and social epistemology, we can logically derive an understanding of what constitutes an epistemological crisis. An epistemic crisis occurs when a group or community, finds itself with reasons to question the correctness of the rules and structures it has traditionally utilized for identifying a true belief.24 In other words, when the traditional methods of discovery that a group or community has relied upon for verifying knowledge are called into question, obfuscated, or disregarded completely, that group can fall into what we can describe as an epistemic crisis. Twin fixtures constitute the crisis itself: doubt about beliefs and doubt about the belief-fixation mechanisms.25 Epistemic crises tend to generally, end in one of two ways: either the crisis is resolved and rational confidence restored in the existing machinery for determining what to believe or, more dramatically, those original prevailing rules and traditions are wholly abandoned and replaced by new principles.26 In order to overcome an epistemic crisis it is imperative for individual epistemic agents within larger epistemic systems to come to consensus on verification systems for knowledge claims that can be widely adopted and accepted. This either involves updating the traditions of epistemic inquiry, creating a whole new set of epistemic principles, or to holding fast to tried and true epistemic traditions in the face of epistemological challenges.

**ALGORITHMS AND EPISTEMIC CRISIS**

The power of algorithms can manifest and be felt in nearly all spheres of society.27 As Frank Pasquale writes, knowing what algorithms do matters, because authority is increasingly expressed algorithmically: “decisions that used to be based upon human reflection are now made automatically.”28 It follows logically, that questions of authority are directly related to questions of epistemology. Noble goes as far as to write that algorithmic systems can structure knowledge and through that structuring, can create a material reality of their own design.29 What
Noble is driving at is that our constant interactions with algorithmically driven systems for nearly all of our information needs begins to manipulate how we perceive our reality. Search results, or social media feeds are sets of algorithmically curated information that can render some important information invisible, some problematic information visible, and can shape our ideas about an alarming number of personal, social and political issues.

In a collaborative study between researchers at Cornell University and Google led by Thorsten Joachims, researchers found that users had a trust bias related to highly ranked items retrieved from search engines, even if those items appeared to be less relevant based on a superficial viewing of the title, abstract, or description of a given item. For more than half of the searches conducted in the study, users failed to look past the fourth result or greater in a Google search. Rather than assessing the actual relevance, users were applying a clearly detrimental epistemological heuristic related to the ranking of a search result in an algorithmically defined hierarchy which has no bearing on that search result’s efficacy or the accuracy of the information contained therein. This is coupled with the notion that algorithmic and other computational tools now play an important political role in areas such as news consumption, issue awareness, and cultural understanding, leading to an increasingly urgent concern within the social sciences regarding how these systems function, influence information dissemination, and how they may be manipulated and perhaps weaponized. Research regarding algorithmic processes within the social media sphere has shown that the algorithms within these systems prioritize sensationalist content that tends to drive user engagement, rather than content that has been fact checked, or evaluated for source authority. Algorithmic systems are performing curatorial duties based on criteria that are unrelated to our traditional epistemic heuristics. There is a mismatch between the public perception of these curatorial criteria, and what the criteria actually are.

An initial example we want to examine to uncover how these algorithmic systems influence our epistemic practices is that of the search algorithm, in this case, Google search in particular. Commercial search tools are inherently trusted, particularly general purpose search engines such as Google, despite being multinational corporations that have effectively privatized information retrieval. The Pew Internet and American Life project noted that as early as 2005, the general public was beginning to see commercial search engines as a form of public institution, similar to the legal, journalistic, and educational institutions we have traditionally relied upon for authoritative information. In a follow up survey in 2012, the same Pew Internet and American Life Project found that almost three quarters of search engine users in the United States relayed that they thought that most or all of the information they find in search engines is accurate and trustworthy, while a further two thirds believe that search engines are a fair and unbiased source of information. It is important to emphasize that there is an imbalance and tension here related to how we perceive the function and role of Google search in our society, and what Google search actually is. Google search seems to be implicitly trusted without any clear evidence that it should be granted such trust and authority.

Google search is essentially an advertising platform, and is not intended to be a public information source. Google's famous “PageRank” algorithm can be generally understood as ranking keyword search results based on popularity which is a subtle manipulation in order to increase engagement and appease advertisers. This is an immediate epistemic red flag as Google search users are clearly misunderstanding the intent and functionality of the search tool itself. The goals of the implementation of the Google search tool are not aligned with the epistemic goals of the general information seeking public, while there is a pervasive assumption by said information seeking public that an alignment of goals does, in fact, exist.

This becomes further complicated when the concept of personalized search based on the collection and analysis of user generated data is considered. Google, as well as almost all other commercial search engines, utilizes tracking scripts in order to monitor user clicks and user responses to system design changes. This data is another aspect of the algorithmic search process in which Google begins to tailor search results based on previous user behavior for the purposes of prioritizing advertising content based upon perceived user preferences. Therefore, Google begins to show search results to users based upon its own sorting and ranking algorithm, which has nothing to do with retrieving the most valuable and credible information based upon a given users keyword input, but in fact, information that is inherently based upon concerns such as popularity, and perceived search user interest based upon collected click and search history data in order to strategically make visible advertisements. Google is essentially predefining a given criteria that is obscured from the user’s view when
choosing to include or exclude search results. This gives agency to an algorithmic search system to mediate how we perceive the given subject of our search. There is little concern within the Google algorithmic ecosystem related to weeding out misinformation, removing hateful and violent content, or addressing misrepresentation of entire sociocultural groups. This phenomenon is particularly troubling in light of well documented problematic cases, such as Noble entering the keywords “black girls” into a google search box and retrieving exclusively pornographic and negatively sexualized representations of black women on the first page of Google search results. Cadwalladr wrote in *The Guardian* about her experience of entering “did the hol….” and being given the subsequent autosuggested search of “did the holocaust happen?”, which resulted in the top ranked search result being titled “Top 10 Reasons Why the Holocaust Didn’t Happen”. In some cases these problematic algorithmic search results have played a pivotal role in the radicalization of troubled individuals such as Dylan Roof, whom discussed his use of Google search in his manifesto. Roof was presented with numerous problematic search results, particularly those published by the white supremacist website of the Council of Conservative Citizens which led him to commit a monstrously violent hate crime, murdering nine human beings in a South Carolina church in June of 2015.

Search results have the power to structure knowledge and create an alternative material reality. Google search is an inflection point for an epistemic crisis. The epistemic opacity of how search results are retrieved and shown to the user, in consort with the mismatch of perception and intent between the users perceived view of the tool and the actual intent behind the implementation and maintenance of said tool is very transparently problematic in epistemic terms. This epistemic confusion results in a complete breakdown of the epistemic process rooted in a misappropriation of authority. User value judgements regarding the actual intent behind the algorithmic search process residing within the black box of Google search obscure the epistemic process of justification based on an ill-informed trust bias.

The other major source of epistemic crisis we will examine is the algorithmic systems found within social media applications, and how those systems can be manipulated to generate epistemic failure both at the individual and social level. If we focus solely on the American context, Facebook is still the most popular social media site, with a large and continuously active user base. According to a pew research report on social media usage from 2014, 70 percent of Facebook users engage with the site on a daily basis, a significant increase over the 63 percent that accessed the site daily in 2013. The sheer scale, and volume of active users and content shared via Facebook is historically unprecedented, and Facebook has fundamentally transformed the ways in which individuals and social groups communicate and build relationships, how they receive and read news, and the ways in which they experience, make sense of, and act upon the world, generally. As the personal becomes political, and increasingly polarized factions of the culture wars in the United States become the most vocal social media users, it is troubling to learn that according to Gottfried and Shearer, fully 60 percent of Americans rely on social media for their discussion of political and social issues. Clearly, the convenience, accessibility and ease of use has provided a great deal of connectivity and given everyone with an internet connection a platform, however, this can lead to some plainly troubling epistemic considerations.

Facebook initially launched the “news feed” in 2006, and have significantly iterated upon that initial launch to include features such as the “like” button, commenting, and also sharing capabilities. The essential functionality of the Facebook news feed algorithm involves the prioritization of “friend” relationships, frequency of interactions with the poster of a story, the number of “likes” and “shares” the post receives, and the interaction history of the user with similar posts. Subsequently, each of these factors are weighted differently based on the type of interaction with both the post itself, and the history of interactions with the user whom made the post, these facets are plugged into a formula, and an affinity score is assigned. In addition to this movement of Facebook posts in or out of the news feed based on user engagement, and connection amongst users, Facebook’s news feed algorithms also push posts higher in the news feed ranking based on the current “trending” status of a topic, or specific item. We can see clearly here that Facebook is constructing a perception of reality outside of the traditional epistemic concerns related to the veracity of information shared. The algorithm itself is using a relatively arbitrarily generated ranking score to pick and choose what to add to and not add to our Facebook news feed, each item is based on an algorithmically perceived preference for a specific form of content at the expense of displaying the entirety of what has been posted within our social network.
A primary epistemic concern is the circularity embedded in the logic of the news feed algorithm. An algorithm that prioritizes not just the general area of interest of a posted item, or when it was posted chronologically, but also the strength of connection between users based on past interactions on the platform clearly creates an epistemic issue related to visibility. For example, research suggests that an average Facebook user should potentially be exposed to 1500+ stories in their news feed upon every instance of a login, however, the news feed algorithm typically whittles this down to around 300 posts. As the algorithm learns more about a Facebook user’s connections to other users in their given network of friends, more data is captured resulting in a news feed that is heavily biased towards the users that one engages with most on the platform. The composition of our friend networks on Facebook is the most important factor limiting the mix of content we encounter. This may have the effect of corralling users and social groups into algorithmically specified information echo chambers. The viewing and consuming of information from a limited subset of individuals that users interact with most often on the social media platform can shape perception as a narrow algorithmically curated conception of reality emerges.

A further epistemic concern related to social media and Facebook in particular is the concept of sharing. Research indicates that Facebook allows a more rapid spread of news through individual sharing of stories. As a result, every individual can become a news source by sharing what they see. This paradigm shift has fundamentally altered both the quantity and quality of information that individuals encounter, and dramatically increases the probability that an individual may encounter, and potentially spread, false information. The epistemology of testimony, a subset of social epistemology, is concerned with discerning the validity of information acquired through others. If a recipient is unable to assess the authority of a teller, or if the results of that assessment suggest fallacy, the recipient may seek others who offer similar testimony to develop a more complete assessment of the credibility of the information. This epistemological phenomenon makes for very murky assignments of authority and justification of belief when we consider that the Facebook news feed algorithm ends up only feeding users a small subset of posts by users that they have interacted with on a regular basis. Justification and verification of information based on the testimony of an algorithmically generated intellectually homogenous echo chamber can end up generating a false confirmation bias of incorrect information.

This phenomenon is particularly troubling as the sophistication of bot technology has increased and has been utilized by political campaigns, state actors, and activist groups in order to manipulate public opinion by disrupting attempts to organize and creating the illusion of popularity to attempt to manufacture false perception of consensus. Studies of process of seeding social media using accounts created by automated bots reveal this technique was utilized to great effect in Michigan in the 2016 United States presidential election, in which the amount of provably false news reports that were shared was at least equal to the amount of professional news in the days leading up to the election. Multiple surveys revealed that a shocking large number of potential voters believed the fake news headlines they were exposed to without attempting to investigate the article itself, or its sourcing further. One particularly absurd piece of fake news that claimed that the Pope Francis of the Roman Catholic Church had endorsed Donald Trump for president of the United States, an easy claim to verify as a falsehood with a quick Google search, was shared or liked on Facebook 960,000 times between its first appearance on Facebook and the 2016 election. Most researchers estimate that this article was viewed, and clicked on, but not shared by a number of users on the order of ten times that number of 960,000. As a point of contrast, the most widely shared article on Facebook throughout the same period of time was a Washington Post piece that was shared only 849,000 times. This is of course, shockingly disruptive to our epistemological practices, particularly when moneyed interests have a stake in shaping and controlling the public narrative. There are examples of this in multiple areas of debate, including the efficacy of vaccines and their potential to cause children harm, the impacts of climate change and subsequently its cause, the negative manipulation of the message at the core of the Black Lives Matter movement and subsequent protests following the murder of George Floyd. Sharing is essential to societal knowledge production and dissemination, but our current information landscape creates an extremely convoluted and densely problematic epistemic path towards justification and verification, two essential components evaluating the veracity of knowledge claims.
CONCLUSION
Algorithmic systems have come to dominate our process of information consumption, and while these systems have provided an unprecedented level of access to the wealth of human knowledge, they have also presented human beings with numerous epistemic challenges. The processes of knowledge justification and verification have been complicated by the opacity of the implementation details of algorithmic systems. The specific functionality of these information systems is poorly understood by the average user, yet they have been granted singular authority over the processes of the selection and dissemination of information. This alone would not be cause for concern, but when we consider research that has shown users to place epistemic trust in these information systems that in most cases are not concerned with the epistemic veracity of the information that is being presented and shared by their algorithmic processes, but are more beholden to fiduciary concerns, the epistemic waters begin to churn and become quite murky and perilous. Epistemic authority and knowledge justification is being misplaced by large numbers of users of these systems, most of whom do not have the time or expertise to adequately discern how the information they are viewing has arrived in their Google search results, or appeared on their Facebook news feeds. Algorithms have the power to structure knowledge and have begun to shape assumptions regarding truth, justification, and authenticity. Without a formidable understanding of the commercial interests shaping the development and implementation of algorithmic systems, there is a level of epistemic opacity that can lead to individual and socio-cultural epistemic crises in which truth and authority themselves are drawn into question. This is deeply troublesome where these systems amplify and preserve marginalizing cultural factors within a society and potentially manufacture consensus regarding negative stereotyping, in turn causing real world harms and continuing the processes of discrimination, prejudice, and marginalization. As academic librarians, particularly in the realm of information literacy instruction, this is of particular import as we must begin to reimagine information literacy instruction in order to include aspects of algorithmic literacy such that we continually ensure that the epistemic practices we are imparting upon our students prepare them to adhere to proper epistemic practices as lifelong learners within an increasingly complex information ecosystem.

NOTES
22. Goldman, 11.
If… Then… Else


27. Bucher, If …Then, 41.
33. Noble, 51.
38. Alex Halavais, Search Engine Society. (UK: Polity, 20017), 47.
39. Fuez, Fuller, & Stadler.
40. Fuez, Fuller, & Stadler.
41. Pasquale, The 58.
42. Noble, 42.
43. Noble, 68.
45. Noble, 155.
46. Noble, 9.
47. Bucher, 66.
49. Bucher, 66.
51. Bucher, 75.
54. Bucher, 81.
55. Bucher, 79.
56. Bucher, 79.
61. Torres, Gerhart, & Negahban, 82.
63. Woolley & Howard, 14.
64. Woolley & Howard, 14.
BIBLIOGRAPHY


O’Connor, Caitlin & Weatherall, 3.

O’Connor & Weatherall, 3.


