Information: Keywords, edited by Michele Kennerly, et al. (Columbia University Press, 2021)

KNOWLEDGE

CHAD WELLMON

In the "Unreasonable Effectiveness of Data," published in 2009, three Google researchers encouraged fellow scholars in natural language processing to forgo "elegant theories," "elaborate models," and complex rules, and to simply follow the data.[1] Given the increasing availability of highly structured data on the web, they suggest scholars interested in designing translation algorithms, for example, should move away from earlier concerns with hand-coded grammars, ontologies, and logic-based expert systems and take advantage of the structure already in the data. Data analysis could replace endless efforts to find linguistic rules and encode them into machines. The exhortations of the Google researchers echoed, although in less exaggerated tones, a similar injunction issued by *Wired* magazine's Chris Anderson a year earlier. Announcing the advent of what he called the "Petabyte Age," he declared that big data and applied mathematics would replace "every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is that the numbers speak for themselves."[2] Humans finally have the tools to mine not just data but knowledge, to extract it from the earth like any other inert resource.

A decade later, predictions that claimed big data will deliver knowledge without theory or causal explanations seem not only naïve but also wrong. For many, these promises turned out to be little more than the most recent return of positivism, a purportedly atheoretical empiricism for our digital age.[3] But they have also raised the prospect of an old dream, reformulated repeatedly from Plato's allegory of the cave to the life philosophy (*Lebensphilosophie*)

of early twentieth-century Germany, of a knowledge closer to the truth of things, one shorn of concepts, models, and theories that get in between the world and the human mind. The desire to escape the cave of shadows, appearances, and corrupted senses and to encounter the real has long motivated not only what and how we know but also why we want to know in the first place.[4]

All the while, machine learning methods and techniques the Google researchers and Anderson touted have become features of everyday life, part of our infrastructures in everything from the recommender algorithms of Netflix to the facial-recognition algorithms of state policing and surveillance.[5] Along with this ubiquity has come more criticism and scrutiny. These more recent debates have focused, however, not on older, standard questions about AI—*what machines can do or still cannot do*—but rather cultural and social questions about justice, equity, and power.[6] Regardless of whether machines can think, the mechanical and computational processes of machine learning can obscure all-too-human biases, prejudices, and power.

In addition to criticisms on its possibly pernicious effects and uses, there are now widespread concerns about the kind of knowledge machine learning may produce. Machine learning, suggest recent critics, produces an unintelligible, possibly inscrutable type of knowledge. Confronted with processes and mechanisms that seem to defy human understanding or whose causal relations cannot be accounted for, scholars are calling for AI and machine learning researchers "to move toward greater transparency and accountability" in how they develop their training data sets and design their algorithms.[7] Concerns such as these make clear that machine learning's purported opacity, not just general ignorance about its techniques and methods, challenges long-standing epistemic ideals, especially the notion of knowledge as justified true belief: the idea that legitimate knowledge can be accounted for and explained by a human knower. This is an ideal with moral weight—real knowledge *ought* to be intelligible. The implication of these calls for greater accountability and transparency is that machine learning would be intelligible if only it were made transparent.

But what do we mean by "transparency and accountability?" And should these ideals orient our notions of what counts as trustworthy knowledge in the twenty-first century? Regardless of whether the wide-spread public interest in machine learning increases or decreases in the coming years, the increased capacities and public scrutiny of machine learning techniques provide an opportunity to reconsider the ideals and commitments underlying predominant and long-standing conceptions of knowledge in western philosophical and cultural traditions—the ideals, norms, practices, and virtues that help determine what counts as knowledge and what is mere information. Such a reconsideration need not necessarily contravene centuries of arguments that relate knowledge to individual mental states and enmesh it with human capacities. But it will show that unintelligibility and inscrutability—as epistemic

anti-ideals—have a history that precedes machine learning. This history can help us better understand the ways in which knowledge—bound up not just in minds but also in media, technologies, practices, and institutions—always exceeds the capacity of any individual mind to possess it and fully account for it. Such a history will not only clarify basic epistemic ideals and norms, such as intelligibility; it will also help us imagine alternatives as we struggle to orient ourselves in our ever-evolving epistemic environments.

***

In the *Meno*, Socrates asks why knowledge is more valuable than "right opinion."[8] Even if right opinions happen to be true, they are not stable. They are like the statues of Daedalus, the ancient Greek craftsman who fashioned sculptures that, as legend had it, could move. Like the statues, right opinions, "are not worth much," Socrates says, "until one ties them down by (giving) an account of the reason why." Knowledge is right opinion that is fastened, grounded in a stable and clear relationship between a person who knows and some given reality or truth. Knowledge, so conceived, entails comprehension, intelligibility, and a level of certainty. In a world filled with the flux of sense impressions, images, and data, true knowledge provides a firm, reliable position. It requires reasons, justification, and, more broadly, a basis for trust.

In Socrates's account, it is the immortal soul's recollection of timeless forms that ties knowledge down, binding that soul with a reality more stable and lasting than any finite body. Although western philosophical traditions have long adopted, adapted, and rejected such a Platonic account, the basic notion that knowledge is primarily a personal and superior mental state has persisted.

From Aristotle to Aquinas, and from Locke to Kant, philosophers have tied real knowledge to individual minds, themselves generally unadorned by technologies and untouched by history. Knowledge, so understood, refers to a capacity to give reasons and to understand why. More contemporary philosophers focus on what they call the "subjective" side of knowledge, seeking to give accounts of the features, properties, and characteristics of this "highly valued state," knowledge, in which an individual person stands in relation to a given reality.[9] Despite continuous disputes and disagreements, key epistemic ideals have remained largely intact. To know, as Descartes put it in 1644, is to hold an idea or perception "very clearly and distinctly."[10] Whereas real knowledge is clear and distinct, false belief, opinion, intuition, or whatever a less-valued form of knowledge might be termed, are fuzzy, opaque, unintelligible—alien to human capacities to account for it.

Given the persistence of these epistemic ideals, what is to be made of deep neural or convolutional networks, algorithms with hidden layers whose outputs and the very steps to produce them are largely incomprehensible?[11] Even though humans have written (or at least copied and modified) the basic lines of code

that constitute such machine learning algorithms and collected the training data upon which they rely, these algorithms combine ever more steps and inputs to produce outputs and behaviors that even their human designers cannot fully account for. It is becoming increasingly difficult, as Thomas Nickles puts it, "to give an account of why" they do what they do.[12]

Contemporary machine learning techniques raise the prospect of a kind of knowledge that cannot be accounted for in the way that Socrates argued was necessary to distinguish knowledge from right opinion. This seems to be the case, in particular, for the outputs of artificial neural networks (ANN), a broad set of widely used computational techniques loosely modeled on the neural structure of the human brain. Neural networks pass inputs (data sets) through a series of layers, each of which consists of processing units called neurons. Most ANNs are made up of three types of layers: input layers which receive the initial data, the hidden layers which extract or filter distinct sets of features from the input layers, and the output layers, which transfer information from the network to the outside world. As the name suggests, the hidden layers have no direct connection with the world outside the neural network. They perform their computations and transformations on the inputs, and thus produce their output from inputs from the neurons of the input layer.

It is the invisibility or hiddenness of these middle layers, where the neural network's filtering and extraction of features happens, that can make the outputs of ANNs seem opaque, inscrutable even to those who might know them best. Some critics might even deny such outputs the honorific "knowledge" and refer to them instead as *mere* information, data, or something else low on the epistemic hierarchy. If an intending knower can neither account nor take responsibility for it, then no human can claim it as her own—justified, stable, clear, distinct—belief. Real knowledge, it would seem, is always personal.

Yet, scholars and intellectuals have long relied on methods, protocols, techniques, media, and technologies to make their encounters and claims of knowledge communicable, visible, repeatable, reproducible, and navigable. Humans rely on tools and technologies not fully in their possession, not fully their own, and, oftentimes, not wholly transparent, in order to justify their opinions—in order to know. Algorithms, including the most complex of artificial neural networks, are the latest tools we use to model and know the world. The barriers to knowledge, then, may lie less with the impossibility of understanding our tools or the inscrutability of our methods and more with the complexity of the world and the finitude of human mental capacities.

From Knowing to Knowledge

Surveying the semantic shifts that "knowledge" (*Wissenschaft*) had undergone over the course of the eighteenth century, a long entry in Johann Adelung's

Dictionary of the German Language, a German multi-volume dictionary first published between 1774–1786 and then in a second edition between 1793–1801, describes a fundamental change in the conception of knowledge. At the beginning of the century, knowledge was used to describe a subjective "condition in which one knows something."[13] By 1800, however, knowledge had come to refer not only to a subjective state but to something objective, something existing beyond any one person— "general truths that were grounded in each other." The first, more "antiquated" definition of knowledge as "particular insight" or mental capacity had given way to another: knowledge as a relationship among ideas themselves and, more broadly, an increasingly distinct realm in which these ideas had taken form—in objects, systems, media, practices, and institutions. Over the course of the nineteenth century, this second notion of knowledge came to predominate, at least in German-speaking lands. Knowledge designated an objective domain that exceeded any one person's capacity to fully possess it. Knowledge was deeply—but not only—human.

In the Adelung entry, in which *Wissenschaft* was presented as the vernacular equivalent of *episteme* and *scientia*, the first definition accorded with a long philosophical tradition of defining knowledge as a personal state of true or justified belief. Consider some of the exemplary images of authoritative knowledge in the European tradition: Descartes's knowing *cogito* who thinks without books and erudition, equipped with only a method and clear and necessary ideas; Locke's individual knower confronting the flux of sense data with nothing but his own mental faculties; or Kant's critical subject who thinks with nothing but naturally endowed categories of understanding and the capacity to synthesize and schematize sensory input.

Each of these images, and the philosophical traditions that sustained and revised them, upheld epistemic self-reliance as the primary epistemic virtue.[14] The ideal of self-reliance and the image of the individual, often heroic, knower who usually accompanied it became acute as philosophers and intellectuals confronted a world they increasingly regarded as bereft of meaningful forms and a divinely guaranteed, rational order. The confidence and hope in a divinely and rationally organized world having waned, the flux and chaos of mere perception and sensory data had to be sifted and organized by human minds.

These shifts in "knowledge" were not simply semantic or philosophical. They point as well to a range of related efforts over the long eighteenth century, born of broader anxieties and anticipations, to reckon with a shared sense of material excess—the proliferation of print as well as observational and, eventually, experimental data. As knowledge came to constitute its own objective reality, scholars and scientists struggled to encounter, engage with, and make sense of an external world saturated with potential knowledge. While some celebrated the growth of this domain as a sure path to intellectual and social progress, others worried that it would soon outstrip human capacities to control and

contain it. They worried about an inevitable gap between two types of knowledge: subjective and objective.

Skeptics described the newly emerging domain of supra-individual knowledge as a distinct world populated by printed things—which, in the second half of the eighteenth century, Johann Gottfried Herder called the "bibliographic Babel" and Novalis the "book world" (*Bücherwelt*)—and warned that it would soon overwhelm individual *cogitos* and minds. In 1750, Rousseau worried that knowledge had begun to outstrip human capacities and comprehension—a lament that would characterize anxieties about technological change to this day, even as the material forms and possible scales of more contemporary digital technologies have introduced new and different possibilities and concerns.[15] Similarly, Rousseau's best student, Immanuel Kant, warned in 1784 that "the book" had come "to think for us."[16] Humans had abdicated their obligation to think for themselves and, as Socrates had predicated in the *Phaedrus*, technical artifacts had come to think for them. The pervasiveness and force of these types of critiques in the final decades of the eighteenth century not only challenged the legitimacy of any extra-individual form of knowledge, they also upheld a distinct, if only implicit, anthropology: the ideal of humans without tools.

This late eighteenth-century anthropology incorporated earlier epistemic ideals rooted in *faculty psychology*, according to which different types of knowledge were ultimately grounded in the unity of the mental faculties. Adapting Bacon's map of learning in *The Advancement of Learning* (1605) and Ephraim Chambers's *Cyclopedia* (2 vols., published in 1728, with 2 supplement vols. in 1753), Diderot and D'Alembert's *Système Figuré*, published in their *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* (1751), mapped the three branches of knowledge to the three human mental faculties (memory, reason, and imagination). Even as knowledge branched out, it was rooted in mental faculties and a shared rational human capacity that allowed all humans to participate in the full flowering of knowledge.

Other scholars and intellectuals, in contrast, embraced the emergent objective domain of knowledge as a secularized space for self-realization, cultural meaning, and human belonging, calling it art, literature, religion—distinct domains of objective and subjective forms of knowledge. Decades before scholars such as Emile Durkheim and Max Weber described western modernity in terms of the differentiation of social spheres, A. W. Schlegel described literature as a distinct and aesthetically superior form of writing; Friedrich Schleiermacher described theology as a particular domain of knowledge about religious experience; and Friedrich August Wolf described philology as a science. Each of these was considered a distinct domain of knowledge with its own traditions, practices, and norms that allowed for human development or *Bildung*. By 1844, Karl Marx could hold up *Wissenschaft* as a space for freedom: "We

must emancipate ourselves," he wrote, "before we can emancipate others." And the path to this freedom lay not in the old oppositions of religion—Christian and Jew—but rather in "critical, scholarly, human relationships [*Verhältnisse*]." Scholarship (*Wissenschaft*) was the "unity" through which the contradictions and illusions of metaphysics, morality, religions, and all other ideologies would be reconciled.[17]

Whether they embraced or feared these objective domains of knowledge, all scholars and intellectuals needed techniques and technologies for navigating, organizing, and searching them. If, as the article in Adelung's *Dictionary* contended, knowledge existed in an objective reality not reducible to individual minds, capacities, or propositions, then its authority and legitimacy could be wholly grounded in individual rational capacities. Furthermore, its transmission exceeded person-to-person exchanges. It had to assume some more public forms, forms that could be assessed and evaluated by a community of knowers. Objective reality was not only that with which an individual knower sought to relate; it was a reality with epistemic potential. As subjective knowledge became objective knowledge, the persona of the distinctly *modern* scholar began to include a capacity to devise and make good use of media through practices and techniques of searching.

Scholars, of course, had long sought to come to terms with the plenitude of information by managing it. Seeking to secure knowledge in the saeculum, early modern scholars developed elaborate note-taking strategies, maintained commonplace books, and formulated reference tools. In her study of how sixteenth-century scholars such as Conrad Gessner and Theodor Zwinger dealt with a prior era of information overload, Ann Blair contrasts information, those "discrete and small-sized items that have been removed from their original contexts," with knowledge, which implies "an independent knower."[18]

It was just this implication of an individual, presumably autonomous knower that nineteenth-century German scholars began to challenge. They transformed common anxieties about overload, surfeit, and proliferation—the ever-increasing material and media of knowledge from periodicals and books to astronomical observations and experimental results—into practical and communal projects for navigating, filtering, and searching the material of knowledge.[19] Scholars' practical need to orient themselves in the ever-expanding domains of knowledge, however, required not just search technologies and techniques but search practices, ideals, and virtues that could form the types of people who could use these tools and better engage with objective knowledge. However complex these domains became, knowledge and knower were never fully severed.[20]

The connections between knowledge and knower were hard won. Scholars developed and cultivated a crucial epistemic ideal: that the objectification of knowledge was also the process of making it common, shared, and

universally communicable. Knowledge was not simply a private good or possession; it was a common and public good as well as an activity. Its creation and transmission required not just individual capacities, insights, and virtues; it also required social practices and virtues that bound individual knowers as scholars working together to sustain collective projects of knowledge.[21] The creation of new epistemic ideals also entailed anti-ideals. For knowledge to be legitimate, it had to be publicly searchable, and therefore could not be private. Knowledge that was not related to other knowledge was not knowledge at all; it was fanaticism, dogma, myth, prejudice—all the *epistemic idols* of modern knowledge.

Yet even as the ideals of the communicability, publicness, and sociality of knowledge became norms, the specter of its incommunicability and opacity remained. The emergence of *knowledge* as a distinct, self-regulating sphere made knowledge more public and accessible, but only for those with access to search technologies and educated in the practices, ideals, and virtues that sustained their right use. It entailed a divide between those who could access these objectified forms of knowledge and those who could not. The habits, cultures, and practices of scholarly and scientific search became key elements of a highly specialized, modern knowledge whose locus was the research university and related institutions that organized and sustained a distinct group of people, practices, and materials.

Yet few, if any, of the scholars and intellectuals who interacted within this objective domain of knowledge would have been able to give step-by-step, rule-based accounts of what they did. The practices, habits, techniques, and cultures that helped constitute knowledge were, in this sense, rarely fully transparent, intelligible, or universally accessible. The epistemic and social value that search tools, techniques, and practices acquired over the course of the late eighteenth and throughout the nineteenth century highlights the limits of publicness, transparency, and intelligibility as epistemic ideals. An unsearchable set of documents, a book not included in a bibliography, or an article with no citations referring to it amount to knowledge that effectively does not exist. What is not part of the whole of knowledge is not really knowledge at all. Whatever remains outside the whole has not been transformed, legitimated, and incorporated into the epistemic ecosystem. Whoever determines or defines the parameters of search—categories, keywords, techniques, and domains—determines what becomes visible as knowledge. Whoever shapes the conditions of access, manages the terms of search, and facilitates the movement of objects in such an environment, helps determine what can emerge as knowledge.

One of the scholars who best articulated and embodied these shifts in the concept and practice of modern, specialized knowledge was the German physiologist Hermann von Helmholtz (1821–1894), who not only made pioneering

discoveries in human physiology and perception but also worked tirelessly to institutionalize knowledge as a collective and shared human enterprise. In 1862, Helmholtz addressed his faculty colleagues at the University of Heidelberg as their newly elected rector and told them that all German scholars faced the same challenge: a profusion of empirical facts. The proliferation of epistemic objects—material things that could be collected, organized, and then marshaled as evidence—had increased as the technologies and techniques for empirical observation had improved.[22] Classical philologists and comparative linguists as well as anatomists and zoologists were so "immersed" in facts that they could not "see anything beyond" the confines of their specialized disciplines. Whether in the form of epigraphic fragments from ancient objects, scattered notes in an archive, or newly collected plant specimens, the sheer stuff of scholarship had begun to make scholars "dizzy" (120).

After Hegel and amidst the rapid expansion of empirical practices and methods, "who," Helmholtz (122) asked his colleagues, would "be able to see the whole," to apprehend the unity of knowledge and maintain it as his personal mental possession? None of them individually, he argued matter-of-factly. There was simply too much to know.[23] In describing how a surfeit of "facts" becomes scholarship, Helmholtz also describes how objective reality (the fact of the world) is distilled into epistemic objects (data and information), which is then transformed into something called knowledge. He lays out a hierarchy according to which individual facts, data, and information are of lesser value than knowledge. For Helmholtz, knowledge (or *Wissenschaft*) is an honorific; it bestows not only a higher value on its referent but also entails norms and ideals about how people ought to regard it or dispose themselves to it.

One of these norms, for Helmholtz as for almost all nineteenth-century German, university-based scholars, was that no one person could account for the totality of knowledge. Helmholtz advised his Heidelberg colleagues to think of knowledge not as something to be held in an individual consciousness but rather as a collective endeavor to be participated in. It was a project sustained by a community of scholars over time. The unity of knowledge was as much an ethical and social project as it was an epistemological one. The task of scholars, philologists and physiologists alike was to develop the means, the media for rendering knowledge communicable across time and space, and to participate in the communities that sustained these media by embedding them in practices and orienting them to common ideals.[24]

In Helmholtz's account, modern knowledge existed in disciplinary domains, or *Fächer*, which balanced well-ordered objective material and well-formed subjective human capacities. Every discipline required both easily accessible and searchable material (lexica, indices, periodicals, encyclopedias) as well as distinctly human capacities (*Geistesfähigkeiten*) that had to be developed and strengthened through repeated exercise. Legitimate knowledge combined both

aspects. The "external" or material organization ensured that even if knowledge could not be readily accounted for or recalled, it "could be found" by anyone at any moment.

Yet, wrote Helmholtz, knowledge could not remain "printed black on white."[25] It had to be taken up, encountered, remade, and transformed by scholars, both individually and collectively. Helmholtz described the knowledge embodied in material forms, from lexica to data sets, as resting in a "field" waiting to be cultivated. His metaphors for knowledge—earth, fields, planting, cultivating, tilling—describe an epistemic ecosystem in which knowledge emerges as the yield of an environment of human, nonhuman, and technological interactions. The task of scholars was to relate the material forms of knowledge—facts, evidence, and observations as transcribed and recorded—to each other and, crucially, to themselves. The material, external order of knowledge, he said, had to be "intellectually conquered."[26] Helmholtz had sketched many of the epistemic ideals that would come to define a distinctly *modern* knowledge: as an endless pursuit; as research; as always changing and constantly being remade; knowledge as never fully intelligible or accountable to any one person.

Google and the Limits of Knowing

How can these historical and theoretical accounts help us better understand what counts as authoritative knowledge today? Although the research universities that Helmholtz upheld as the key institutions of nineteenth-century knowledge continue to play a crucial role in our current epistemic and media environment, they increasingly do so alongside, or even at odds with, digital platforms and corporations, such as Google. Even as the trust and confidence of people across the globe in media, politicians, and universities steadily erodes as populist protests have grown, Americans, at least, continue to trust their search results.[27] And yet, Google's search algorithms remain fundamentally inscrutable, even if their training logics and search results are not.[28]

If eighteenth- and early nineteenth-century readers faced a surfeit of print, computer engineers and early users of the World Wide Web Project in the early and mid-1990s faced an exponential increase in the number of webpages. Two years after the World Wide Web Project began in 1991, there were only 130 websites. By the time Yahoo was founded in 1994, there were 2,739 websites. Four years later in 1998 when Google was founded, there were around 2,410,067 websites, and just two years later in 2000 over 17,000,000 websites. Today, in 2018, there are over 1.85 billion websites.[29]

The rapid growth of the WWW presented big challenges to the methods of early search engine companies—such as Lycos, Infoseek, AltaVista, and Yahoo—that were using automated crawlers to follow links, copy the pages, store them in an index, and then use human labor to create lists of keywords and associated websites. In their original paper outlining the "anatomy of a large scale hyper-textual Web search engine," Page and Brin proposed a different way of approaching the problem of search. They began from the insight that the web "was loosely based on the premise of citation and annotation—after all what is a link but a citation and what was the text describing that link but annotation."[30] They sought to create a model of the citational structure of the web as constituted by links among pages, and eventually developed a proprietary algorithm that modeled the links, not only the outgoing ones but also their backward paths, that constituted the web: PageRank.

The crucial distinction between Google PageRank and these first-generation web search engines was that Brin and Page had argued that the quality of a page was a function of its position within the network of webpages. What made a piece of information valuable was not the class or category to which it might belong, but rather the relationships it had to other pieces of information.

There is an important continuity between the print techniques and technologies developed over the eighteenth and nineteenth centuries and Google's early attempts "to organize the world's information." Both projects were premised on the idea that knowledge exceeded any personal mental state; it was presumed to exist independent of any one individual, embodied in printed objects or digital structures. For its first decade, Google's leaders and engineers imagined the world wide web and digital forms of knowledge in terms of print. Like their print predecessors who sought to organize the "world of books," Brin and Page sought to "brin[g] order to the web."[31]

And yet, Google PageRank can search only that which has already been linked to the web; its results are entirely imminent to the web's structure. And so, it values only that which has already been valued, that is, what has been linked to by other web pages. Because PageRank models the web, there will always be gaps in Google-knowledge. An unindexed website cannot be searched and, thus, given Google's near monopoly on search tools for the web, essentially does not exist.

Furthermore, the parameters of PageRank—every tweak, every adjustment, every added parameter to its basic algorithm—determine, in conjunction with any given search term, what websites are returned and their rank, and which ones are not. Just as those who defined and managed the parameters of nineteenth-century search technologies, those who manage Google's search engine help determine what counts as knowledge. These interventions are based on human decisions and actions that are rarely made public and are definitely

not subject to public deliberation. They are the decisions of a corporation driven by capital interests.

In its first decade, Google showed little interest in content webpages. According to Google, epistemic authority or legitimacy was simply a function of the citation (link) graph of the web—the authority of a website corresponded to its popularity. Over the past decade, however, Google engineers and executives have gradually begun to discuss fundamental changes not only to its search algorithms but also to Google's evolving epistemic ideals. Google seems intent on becoming not just the organizer of "information" but the arbiter of knowledge. As one of its engineers blogged in 2012, Google was transforming itself from an "information engine" to a "knowledge engine." Frustrated by the "document-centric" character of PageRank, Google has recently sought to develop search technology that "liberates" data from documents and uses that data to create knowledge. One of its first public projects to attempt the creation of knowledge was Knol, a now defunct effort to establish a Google Wikipedia, an online encyclopedia of individually authored articles and essays.[32] Whereas PageRank legitimates a webpage by evaluating its position in the link network of the web, Knol legitimated a page by relating it to a particular person. Knol was based, that is, on a more traditional form of epistemic authority: people as worthy of trust and, thus, reliable sources upon which to justify beliefs. Such a belief is based on the reliability of a known author or authority, whose evidence and arguments can be tested and evaluated.

Although Google ultimately abandoned Knol, deleting it in 2012, the company continues to pursue the creation of knowledge over the *mere* organization of information. In a research paper published in 2015, a team of Google engineers presented a new search method that relies not on "exogenous signals" (links) but on "endogenous" ones (facts). In extracting "facts" and then evaluating websites based on the "correctness" of these facts, Google's engineers are attempting to determine the value or authority of a website based on factors or characteristics not imminent to the link graph structure of the web but on things given—facts—external to that structure. Such a process yields a trustworthiness score or, in Google talk, a *knowledge-based-trust* (KBT), that defines trustworthiness as the probability that a web source contains the correct "value for a fact."[33] This probability is largely determined by comparing an extracted "fact" to potentially similar ones collected in separate (Google owned) databases. Like Socrates's imperative to control the statues of Daedalus, Google is trying to tie down its knowledge by stabilizing its facts. But instead of tying knowledge to an immortal soul, a community of researchers, or a textual tradition, Google's engineers are tying it down to its ever-expanding collection of databases. Google's interest in *trustworthiness* exemplifies its effort not just to organize but to redefine what counts as knowledge. Instead of simply modeling the web's inherent link structure, so redolent of

eighteenth- and nineteenth-century indexical print technologies, Google hopes that it might one day, as Brin put it, "understand," that it might in some sense "know."

Google's desire to transcend the document-centric web is a desire to liberate knowledge from the stubborn particularity of pages of texts and transcend the history of knowledge as the interaction of media, people, institutions, and practices, and not simply a subjective state or inert object. In a perhaps ironic historical twist in the history of knowledge, Google engineers are seeking a way of knowing that is purportedly less susceptible to the manipulations and desires of others, a way of knowing that is more stable and reliable. But what norms, practices, and values would orient this ostensible liberation of knowledge from texts? Who, as Helmholtz asked, sets the parameters and ends of search in a post-link epistemic environment?

***

Knowledge, writes the philosopher Linda Zagzebski, is "cognitive contact with reality."[34] Although philosophers have long focused on accounting for how such contact is possible through individual acts of intellection, any account of knowledge must also consider how such contact also requires complex relations of individual and communal or shared actions, capacities, and habits as well as their objects, technologies, and techniques. Conceiving of knowledge in these more environmental terms can help us better understand how knowledge becomes communicable, sharable, and, in some way, a common possession, not simply a personal state or belief. Doing so can also alert us to the ways in which knowledge is regulated, guarded, and controlled. Ultimately, it can help us understand knowledge as an emergent element of an epistemic ecosystem, in which the material objects of knowledge and the activities and people associated with them are coordinated.[35] Knowledge, so conceived, is not a property or evaluation of any one element—the status of a personal belief or the content of a text—but rather a good born of complex relations not always immediately intelligible and sometimes even inscrutable.

The prospect that machine learning might introduce a knowledge wholly inscrutable and alien to humans is an opportunity to reconsider our assumptions about reason, rationality, and knowledge.[36] Perhaps it is not only the inner layers of neural nets that are unintelligible, but also the norms governing how we learn, know, and orient ourselves in the world.

We need to understand better how knowledge, especially in our digital age, exceeds any individual person's capacity to justify a particular belief. The idealized individual knower, the figure of the autonomous epistemic subject—justified in her belief and capable of accounting for it—limits our understanding of the conditions of legitimate, authoritative knowledge. So too does

the tendency among some media theorists to dismiss human capacities and distinctly human concerns and cares as vestiges of a romantic (or humanist or religious) ideology. Humans, both individually and as collectives, act and think in the world using their technologies, but they are not reducible to those technologies. The authority and legitimacy of knowledge is bound up not only with its material media but also with the character, capacities, and virtues of knowers who make their way in the world by means of these media.

Notes

1. Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems* (2009): 8–12.
2. Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23, 2008.
3. Matthew L. Jones, "How We Became Instrumentalists (Again): Data Positivism Since World War II," *Historical Studies in the Natural Sciences* 48, no. 5 (November 2018): 673–84.
4. L. M. Sacasas, "The Allegory of the Cave for the Digital Age," The Frailest Thing, December 1, 2018, https://thefrailestthing.com/2018/12/01/the-allegory-of-the-cave-for-the-digital-age/.
5. For a largely enthusiastic account of the ubiquity of machine learning techniques, see Klaus Schwab, *The Fourth Industrial Revolution: What It Means and How to Respond,* https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab.
6. Hubert Dreyfus, *What Computers Still Can't Do* (Cambridge, MA: MIT Press, 1972). For the more recent critiques, see, for example, Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge: Polity, 2019); Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).
7. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, et al., "Datasheets for Datasets," Preprint (March 23, 2018). See also Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, et al., "AI4People—an Ethical Framework for a Good AI Society: Opportunities, Principles, and Recommendations," *Minds and Machines*, November 26, 2018.
8. Plato, *Five Dialogues*, trans. G. M. A. Grube (Cambridge: Hackett, 1981), 85.
9. Linda Zagzebski, "What Is Knowledge?" in *The Blackwell Guide to Epistemology*, ed. John Greco and Ernest Sosa (Oxford: Blackwell, 1999), 92–116.
10. Descartes, *Meditations on First Philosophy,* in *The Philosophical Writings of Descartes*, vol. 2, trans. John Cottingham (Cambridge: Cambridge University Press, 2008), 1–62.
11. John Brockman, ed., *What to Think About Machines That Think* (New York: Harper Perennial, 2015).
12. Thomas Nickles, "Alien Reasoning: Is a Major Change in Scientific Research Underway?" *Topoi* (2018).
13. Johann Christoph Adelung, *Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart* (Vienna: Bauer, 1811), 4:1582–83. For an extended discussion, see Chad Wellmon, *Organizing Enlightenment: Information Overload and the Invention of the Modern Research University* (Baltimore: Johns Hopkins University Press, 2015), 37–38.

14. See Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (New York: Oxford University Press, 2012).
15. Jean-Jacques Rousseau, *Discourse on the Sciences and the Arts, First Discourse*, in *The Social Contract and Discourses*, trans. G. D. H. Cole (London: Everyman, 1993).
16. Immanuel Kant, "Beantwortung der Frage: Was ist Aufklärung?" in *Gesammelte Schriften*, ed. the Königliche-preußischen Akademie der Wissenschaften, 29 vols. to date (Berlin: de Gruyter, 1902), 8:35.
17. Karl Marx and Friedrich Engels, Werke, "Zur Judenfrage" (Berlin: Karl Dietz Verlag, 1976), Band 1, 347–77, 349.
18. Ann Blair, *Too Much to Know: Managing Scholarly Information Before the Modern Age* (New Haven, CT: Yale University Press, 2010), 2.
19. Gygory Markus, "Changing Images of Science," *Thesis Eleven* 33, no. 1 (August 1992): 1–56.
20. Lorraine Daston and Peter Galison, *Objectivity* (New York: Zone, 2007), 375.
21. See Alex Csiszar, *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century* (Chicago: University of Chicago Press, 2018).
22. Hermann von Helmholtz, "Über das Verhältniss der Naturwissenschaften zur Gesammtheit der Wissenschaft," in *Vorträge und Reden*, 3d ed. (Brunswick: Friedrich Vieweg, 1884), 120, 122. Further citations included in the text.
23. For an important account of an earlier moment of felt excess, see Blair, *Too Much to Know*.
24. Lorraine Daston and Peter Galison, *Objectivity* (New York: Verso, 2010), 255.
25. Helmholtz, "Über das Verhältniss," 124.
26. Helmholtz, "Über das Verhältniss," 124.
27. Adam Epstein, "People Trust Google For Their News More Than Actual News," *Quartz*, January 18, 2016. https://qz.com/596956/people-trust-google-for-their-news-more-than-the-actual-news/.
28. See, for example, Noble, *Algorithms of Oppression*.
29. "Total Number of Websites," *Internet Live Stats*, http://www.internetlivestats.com/total-number-of-websites/#trend.
30. John Batelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture* (New York: Portfolio, 2005), 72.
31. Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* 30 (1998): 107–77.
32. Udi Manber, "Encouraging People to Contribute," *Google Official Blog*, December 13, 2007, https://googleblog.blogspot.com/2007/12/encouraging-people-to-contribute.html.
33. Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, et al., "Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources," *Proceedings of the VLDB Endowment* (2015).
34. Zagzebski, "What Is Knowledge," 92–116, 109.
35. Alex Csiszar, "Seriality and the Search for Order: Scientific Print and Its Problems During the Late Nineteenth Century," *History of Science* 48, nos. 3/4 (2010): 399–434.
36. For an accessible account of this concern and the formulation of "alien knowledge," see David Weinberger, "Our Machines Have Knowledge We'll Never Understand," *Wired*, April 17, 2018, https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/.