

Exploration and Exploitation: An Essay on (Machine) Learning, Algorithms, and Information Provision

*Deven R. Desai**

Legal and regulatory understandings of information provision miss the importance of the exploration-exploitation dynamic. This Essay argues that is a mistake and seeks to bring this perspective to the debate about information provision and competition. A general, ongoing problem for an individual or an organization is whether to stay with a familiar solution to a problem or try new options that may yield better results. Work in organizational learning describes this problem as the exploration-exploitation dilemma. Understanding and addressing that dilemma has become a key part of an algorithmic approach to computation, machine learning, as it is applied to information provision. In simplest terms, even if one achieves success with one path, failure to try new options means one will be stuck in a local equilibrium while others find paths that yield better results and displace one's original success. This dynamic indicates that an information provider has to provide new options and information to users, because a provider must learn and adapt to users' changing interests in both the type of information they desire and how they wish to interact with information.

Put differently, persistent concerns about the way in which news reaches users (the so-called "filter bubble" concern) and the way in which online shopping information is found (a competition concern) can be understood as market failures regarding information provision. The desire seems to be to ensure that new information reaches people, because that increases the potential for new ideas, new choices, and

* Associate Professor of Law and Ethics, Georgia Institute of Technology, Scheller College of Business, J.D., Yale Law School; former Academic Research Counsel, Google, Inc. This Essay has benefitted from discussions with and input from Danielle Citron, Brett Frischmann, James Grimmelman, and Peter Swire. This Essay was supported in part by summer research funding from the Scheller College of Business and an unrestricted gift to the Georgia Tech Foundation, Inc. by Google, Inc. The views expressed herein are those of the author alone and do not necessarily reflect the view of those who helped with and supported this work.

new action. Although these desired outcomes are good, current criticisms and related potential solutions misunderstand the nature of information users and especially information provision, and miss an important point. Both information users and providers sort and filter as a way to enable better learning, and learning is an ongoing process that requires continual changes to succeed. From an exploration-exploitation perspective, a user or an incumbent may remain isolated or offer the same information provision but neither will learn. In that case, whatever short-term success either enjoys is likely to face leapfrogging by those who experiment through exploration and exploitation.

INTRODUCTION	543
I. THE HISTORY OF ORGANIZING INFORMATION	546
A. <i>Past and Present Concerns About Information Provision</i>	547
B. <i>Digital Approaches to Organizing Information</i>	550
II. THE INFORMATION USER: HOW INDIVIDUATION CAN FEED	
INDIVIDUALITY	557
A. <i>The Quest for Relevancy and Maybe Serendipity</i>	559
B. <i>User Tensions: Isolation, Connection, and Action</i>	562
III. THE INFORMATION PROVIDER: THE NEED FOR EXPLORATION	
AND EXPLOITATION	567
A. <i>Learning Through Exploration and Exploitation</i>	568
B. <i>Information Provider Tensions: Desires for Walled</i>	
<i>Gardens</i>	573
C. <i>Open Networks and Popping Bubbles</i>	576
CONCLUSION: IMPLICATIONS AND LIMITS	578

The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships.¹

-Vannevar Bush, 1945

1. Vannevar Bush, *As We May Think*, 4 ACM INTERACTIONS, Mar. 1996, at 36, 37 (1979) (reprint of the original 1945 article that appeared in the *Atlantic*).

*One of the most important problems in machine learning—and life—is the exploration-exploitation dilemma. If you’ve found something that works, should you just keep doing it? Or is it better to try new things, knowing it could be a waste of time but also might lead to a better solution?*²

-Pedro Domingos, 2015

INTRODUCTION

Recent questions over information organization and provision and competition in those efforts miss a perspective: that of the information provider. Here we are talking about a provider as an entity that aids a user in finding information yet also makes choices about how to organize and display that information.³ Concerns about how an information provider displays results involve two claimed problems. One asserted problem is that an information seeker, or user, may employ, or be subject to, information provision tools that filter results so that he or she sees only information that he or she likes. Thus, a conservative or a liberal or anyone with a particular view would only see news with which they agree. Such a user arguably ends up living in a filter bubble or an echo chamber.⁴ Another claimed problem is that as information providers expand their businesses, other businesses that wish to reach users will be unable to do so.⁵ The fear is that as Apple, Amazon, Facebook, Google, Microsoft, Pinterest, Twitter, Yelp, and so on offer more than their initial services and digital services converge,

2. PEDRO DOMINGOS, *THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD* 129 (2015).

3. See, e.g., James Grimmelman, *Speech Engines*, 98 MINN. L. REV. 868, 880-889 (2014) (describing the differences between the conduit view of information intermediaries, that seeks to treat such information connectors as neutral map makers, directors of traffic, and limit possible, perceived bias, as opposed to the editor view that sees such intermediaries as choosing amongst options to offer what the editor sees as the best result). Here I use the term to be a broader notion than the legal definition in Section 230 of the Communications Decency Act, which follows the conduit view. See 47 U.S.C. § 230(C)(1) (2006); accord Grimmelman, *supra* at 944 (noting the conduit theory view of Section 230). I thank Danielle Citron for pressing me to clarify this point.

4. See, e.g., Natalie Jomini Stroud & Ashley Muddiman, *Exposure to News and Diverse Views in the Internet Age*, 8 I/S 605, 616 (2013) (noting the concern over possible echo chambers); Dan Hunter, *Philpic.com*, 90 CALIF. L. REV. 611, 615 n.17 (2002) (reviewing Cass Sunstein’s book *Republic.com*).

5. See, e.g., Press Release, European Commission, Antitrust: Commission Sends Statement of Objections to Google on Comparison Shopping Service (Apr. 15, 2015), http://europa.eu/rapid/press-release_MEMO-15-4781_en.htm (expressing “concern” that Google’s shopping service operates “to the detriment of consumers and rival comparison shopping services, as well as stifling innovation”).

users will not know about, let alone use, competitors' offerings.⁶ In simple terms, we might see these concerns as identifying a type of market failure. I argue that the essence of both these claims can be understood as one of making sure that new information reaches people, because that increases the potential for new ideas, new choices, and new action. Although these desired outcomes are good, current criticisms and related potential solutions misunderstand the nature of information seeking and especially *information provision*, and miss important points. Both information seekers and providers sort and filter as a way to enable better learning, and learning is an ongoing process that requires continual changes to succeed.

Vannevar Bush's 1945 article "As We May Think" captured why we sort and filter information and how that enables learning.⁷ Even seventy years ago there was too much information to navigate. Then, as now, a common approach to coping with large amounts of information—what Bush called the "record"—was to index it.⁸ But as Bush pointed out, indexing only goes so far in helping us "make real use of the record."⁹ The problem was and is that the record of information continues to grow, but "we can hardly consult it."¹⁰ Although selection and indexing can help us consult the record, "[o]ur ineptitude in getting at the record is largely caused by the artificiality of systems of indexing and storage."¹¹ The quest to improve information provision is in part a quest to improve any one of our abilities to find that which we seek. As such, this Essay offers that a better understanding of the history of information sorting, recent studies on how people engage with information online, the exploration-exploitation dynamic within machine learning approaches to information provision, and the nature of networks shows that the problem today is not alleged filter bubbles or unfairness in providing information.¹² I argue that as before, and perhaps always with information, the problem is about further

6. Cf. Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries*, 104 NW. U. L. REV. 1, 119–23 (2010) (discussing ways intermediaries such as broadband and media companies together might manipulate access to information without users's knowledge of such acts).

7. See generally Bush, *supra* note 1.

8. *Id.* at 37.

9. *Id.*

10. *Id.* at 42.

11. *Id.* at 43.

12. For example, James Grimmelmänn's recent work on speech and search provides a speech-based analysis of search provider's rights and touches on filter bubble arguments. See generally, Grimmelmänn, *supra* note 3. The work explains the tensions in the legal perspectives on search quite well, but I disagree with his normative claims.

increasing everyone's "*present ability to make real use of the record.*" Whether that ability is present and what it means for each of us to make full use of the record requires a fuller understanding of the nature of information seeking and provision.

I begin in Part I with a history of how we sort information in both the analog and digital world. I show that as soon as one seeks relevance, the question of what is the best result—that part of the record that a specific person wishes to use—has political implications and is often "inherently subjective" as that phrase is understood within information science.¹³ In Part II, I turn to the information user and the tensions between individuation—specific identification of someone—and protecting individuality. This tension relates to questions of relevancy, serendipity, and whether online filters create a world of people shielded from views they dislike and who in turn forgo political engagement or are prevented from taking new actions.

In Part III, I examine the information provider perspective and especially issues of exploration and exploitation as they relate to learning. Drawing on organizational learning and computer science literature on machine learning, I show that given the number of users and their subjective desires combined with the ever-growing amount of information and other information providers in the marketplace, a provider must do two things and repeat them. A provider must first explore to learn a strategy about how to proceed. It is through learning that a provider can devise a good way forward. After that, a provider must exploit what it has learned—i.e., work within what is known as a specific optimal strategy—to meet a specific user's interest, which is a filtering process. At the same time, a provider must continue to explore and thus vary results to see whether it is matching a user's subjective needs. This dynamic is vital to any information provider as a matter of business strategy and viability lest others learn better strategies and leapfrog the previously successful incumbent's strategy. Specifically, the dynamic requires at least some variance in what is shown to a seeker of news, shopping, or other information. I also examine how open networks hinder, and perhaps prevent, the ability to create filter bubbles and walled gardens that thwart competition. I conclude with some observations about the noneconomic issues that may lie behind the

13. This discussion draws on and owes much to the so-called "search neutrality" literature. For an excellent summary of the problems with the search neutrality approach and its flaws, see James Grimmelman, *Some Skepticism About Search Neutrality*, in *THE NEXT DIGITAL DECADE: ESSAYS ON THE FUTURE OF THE INTERNET* 435, 438 (Berin Szoka & Adam Marcus eds., 2010) (listing eight potential "search-neutrality principles" and finding all "unusable as bases for sound search regulation").

current debate over information provision, competition, and the use of algorithms, as I identify what lessons those issues offer as a caution to incumbent information providers.

In short, by understanding the dynamics of how users seek and how information providers organize and provide news as a type of information, we will see whether filter bubbles exist, and if they do, what works to prevent them from being stable. That investigation will help see that concerns over competition operate under similar rules. If news in general can and does reach people despite overt filtering possibilities, it seems unlikely that news of shopping would somehow not be found. Nonetheless, these questions may offer guidance to information providers about other values and larger, underlying issues that explain the recurrent concerns about news, online competition, and information provision in general.

I. THE HISTORY OF ORGANIZING INFORMATION

The history of organizing information can be understood as the move from indexes to approximating how the mind works. Categorizing and organizing allows us to manage information.¹⁴ As the volume of information grows, we all seek ways to parse that information. The point of this effort is not to know or see all that is knowable or seeable, but to know how to find, think about, and use that piece of the information world that connects to one's inquiry and needs.¹⁵ The simplest way to organize information is by indexing. As one computer scientist has said, "search engines did not invent indexes: in fact, the idea of indexing is almost as old as writing itself."¹⁶ Universities, libraries, encyclopedias, thesauri, syllabi, and more have been with us for centuries as ways to filter the vast amounts of knowledge the analog world offered.¹⁷ Today, those methods have a place, but the scale of

14. See, e.g., JOHN MACCORMICK, *NINE ALGORITHMS THAT CHANGED THE FUTURE: THE INGENIOUS IDEAS THAT DRIVE TODAY'S COMPUTERS* 12 (2011) (eBook) ("For example, archaeologists have discovered a 5000-year-old Babylonian temple library that cataloged its cuneiform tablets by subject. So indexing has a pretty good claim to being the oldest useful idea in computer science.").

15. Cf. FRANK WEBSTER, *THEORIES OF THE INFORMATION SOCIETY* 28–29 (John Urry ed., 3d ed. 1995) (noting the difference between the amount of information and "the meaning and quality of the information").

16. See, e.g., MACCORMICK, *supra* note 14, at 12.

17. See, e.g., DAVID EASLEY & JON KLEINBERG, *NETWORKS, CROWDS, AND MARKETS: REASONING ABOUT A HIGHLY CONNECTED WORLD* 382 (2010) ("[I]nformation networks date back into much earlier periods in our history; for centuries, they were associated with libraries and scholarly literature, rather than with computer technology and the Internet."); RICHARD LANHAM, *THE ECONOMICS OF ATTENTION: STYLE AND SUBSTANCE IN THE AGE OF INFORMATION* 13–14 (2006).

information is greater. Algorithms, public experts, social networks, online rating systems, and more have emerged to help us as we again try to sort information overload.¹⁸ As the amount of information grows, we may stand in awe of the quantity but we can say little more than there is more.¹⁹ In the quantitative approach, all information is treated the same regardless of semantic meaning.²⁰ That approach has power and use in some contexts and as long as one need not distinguish amongst information. Once one asks whether information is “significant, accurate, absurd, interesting, adequate or helpful”—one might say relevant—things change.²¹

A. Past and Present Concerns About Information Provision

We often like and seek indexes and filters, but it is easy to forget that they almost always involve choices and sometimes politics by both users and organizers. For example, the organizing pillars of academic life seem neutral, but one important and necessary thing teachers do is organize and select information to put it into coherent systems.²² After that effort, we can draw on their expertise and add it to our own views.²³

Even in the limited context of a specific academic discipline there are many options for an information seeker. If you want to learn about history, biology, economics, environmental science, or any other subject, a good academic article or book can help. The author will have read a range of materials, decided which parts to include, and organized the bits and pieces into a coherent structure on which you and others can draw.²⁴ If you want direct access to the sources used in the publication,

18. See, e.g., Frank Pasquale, *Copyright in an Era of Information Overload: Toward the Privileging of Categorizers*, 60 VAND. L. REV. 135, 136–37 (2007) (praising the rise of tastemakers who use metadata because metadata is “essential to finding the expression one wants” and celebrating a world where “more and more services rate and organize content, [such that] there is less reason to think one has missed some particularly compelling, delightful, or important work”); cf. LANHAM, *supra* note 17, at 17 (arguing that web designers work to direct and capture attention on the Internet).

19. WEBSTER, *supra* note 15, at 27.

20. This move is Claude Shannon’s breakthrough in information theory related to communicating a message through a channel. See C.E. Shannon, *A Mathematical Theory of Communication*, 27 BELL SYS. TECH. J. 379 (1948).

21. WEBSTER, *supra* note 15, at 27; cf. MACCORMICK, *supra* note 14, at 17 (“What does the ‘rank’ of a page really depend on? The real question is not ‘Does this page match the query?’ but rather ‘Is this page relevant to the query?’”).

22. See, e.g., LANHAM, *supra* note 17, at 13–14.

23. *Id.*

24. UMBERTO ECO, *HOW TO WRITE A THESIS*, 6, 11–13 (2015) (explaining that a good work will have “[i]dentified a precise topic,” “[c]ollected documents on that topic,” “[o]rdered these documents,” and “[p]rovided the necessary documentation so that readers may reexamine the topic through his sources”).

you could check the citations and bibliography (both are indexes of a sort), go to a bookstore and buy the books or go online and search for cited articles. If you want a smaller set of materials, you could, as I often do, go to a university bookstore to see what books have been assigned for a class.

Yet the course itself, the structure of the reading and the articles, is not well captured by such a trip. You would not know in what order to read the books, whether only part of a book was assigned, and what supplemental material was listed in the course syllabus and reading list. Offerings such as MIT's OpenCourseWare project now bridge this gap.²⁵ The material is still organized as it would be in analog settings, but by being digitized and placed online, anyone with an Internet connection can draw on an expert's view of the basic, intermediate, and advanced materials. The professor has filtered the texts and given a different roadmap to the field than the author of a given book.²⁶ Thus, there are several different ways to understand what is relevant *and* to organize whatever information is deemed relevant.

Although academic organization of information appears neutral, reflection on the so-called culture wars shows that education, especially how it is organized, can be political.²⁷ For example, the curriculum framework for the AP U.S. History test has been contested for the past few years.²⁸ This debate can be seen as part of an ongoing concern about education and what is best for students, something that began at the least with Allan Bloom's *The Closing of the American Mind*.²⁹ In other words, debates about information provision even in this limited setting reveal issues that have been present for some time. Yet this

25. MIT OPENCOURSEWARE, <http://ocw.mit.edu/index.htm> (last visited Nov. 24, 2015).

26. If one wants to take the course, companies such as Coursera, EdX, and Udacity allow that too.

27. See, e.g., Anya Kamenetz, *The New, New Framework for AP U.S. History*, NPR ED (Aug. 5, 2015, 5:17 AM), <http://www.npr.org/sections/ed/2015/08/05/429361628/the-new-new-frame-work-for-ap-u-s-history> (explaining the fight over the 2014 curriculum framework for the AP U.S. History test, a recent example of the political nature of education).

28. The Republican National Committee, among others, opposed the changes as being "radically revisionist" and focusing on ideological views rather than facts. *Id.*

29. ALLAN BLOOM, *THE CLOSING OF THE AMERICAN MIND* (1987); see Donald Lazere, 'The Closing of the American Mind,' 20 Years Later, *INSIDE HIGHER ED* (Sept. 18, 2007), <https://www.insidehighered.com/views/2007/09/18/lazere> (offering that Bloom's book has been considered "the opening salvo in a ceaseless conservative war against the academic and cultural left"); see also Wilfred McClay, *The Legacy of "The Closing of the American Mind,"* *IMAGINATIVE CONSERVATIVE* (May 20, 2015), <http://www.theimaginativeconservative.org/2015/05/legacy-closing-of-the-american-mind.html> ("The book's arrival corresponded, as did the culture wars themselves, with a noticeable deepening of American academia's intellectual and moral crises.").

example differs from media concerns, in part because the changes in AP and other curricula is public and arguably subject to debate and reform.³⁰ Thus there have been aspects of a given medium's role in information provision that are related to, but expand, the critiques about information selection and provision.

Nonetheless, the general concern that provision of information via a new medium possibly harms public life by turning people inward and limiting action is not new.³¹ Newspapers, publishing houses, radio and television stations and networks, bookstores, and many other purveyors of information performed filter functions in the past and still do so today. As far back as 1933—and throughout the decades since—theorists have been concerned that information power and new media supported consumer capitalism over all else.³² That way of life was thought to be “home-centered to the detriment of civic relations.”³³ The claim and fear was that “people [would be] predominantly passive” and “hedonism and self-engrossment predominate and [would] find encouragement” while “public virtues such as neighborliness, responsibility, and social concern” languish.³⁴ Today, those ideas have been resurrected on the theory that because information media are now digital, there is a new urgency about sorting, organizing, and controlling what people see, read, and hear.

Search, social networks, online rating systems, tweets, apps, and mobile computing have emerged to aid us as we try to make sense of the world, and these advances generate the perceived problem of perfect filtering, echo chambers, and walled gardens. As with other media, there is a history to these claims. More than ten years ago, Dan Hunter explained that the fear of perfect filtering and related filter bubble concerns were not at hand.³⁵ Cass Sunstein, a proponent of filter bubble logic, tried to claim the idea was more of a thought experiment.³⁶ Nonetheless, with improved technology and changes in online services, Sunstein and others argue that these changes have finally achieved dangerous levels of filtering that have a deadening effect on political action by allowing us to avoid that which is uncomfortable.³⁷ Recent

30. I thank Danielle Citron for calling out this distinction.

31. See, e.g., WEBSTER, *supra* note 15, at 94–95.

32. *Id.*

33. *Id.* at 95.

34. *Id.*

35. Hunter, *supra* note 4, at 614.

36. See CASS SUNSTEIN, *REPUBLIC.COM* 208–09 (2001).

37. CASS R. SUNSTEIN, *REPUBLIC.COM 2.0*, at 148, 220 (2007); see ELI PARISER, *THE FILTER BUBBLE* 30 (2011) (noting that although people can actively chose information regarding a particular political viewpoint through news sources, people do not “chose to enter the filter

antitrust investigations in the EU on how Google displays shopping results operate from a similar logic: that users are prevented from finding new options, or ones differing from a provider's interests, online and as such society is harmed.³⁸ In the European Commission's words, the claim is "that users do not necessarily see the most relevant results in response to queries."³⁹

In both the online and offline worlds, information is sought, and information is organized. Seekers desire guidance, and organizers hope to provide it. Seekers have options, and organizers offer different paths to knowledge. In both arenas, people debate what is relevant and politics can underlie concerns about what information is offered and what is left out. Although choices and politics affect information provision in general, we need to understand the differences that arise with digital approaches to information provision to understand today's concerns.

B. Digital Approaches to Organizing Information

Although we can see that organizing information involves choices and politics, there is a certain romanticism of the objective information provider at work in discussions of digital approaches to organizing information.⁴⁰ The history of directories and search—including Silicon Valley hubris during that history—deserves partial blame.⁴¹ That

bubble" of Google).

38. See, e.g., *supra* note 5.

39. *Id.*

40. See, e.g., *The Power of Google: Serving Consumers or Threatening Competition?: Hearing Before the S. Subcomm. on Antitrust, Competition Policy and Consumer Rights*, 112th Cong. 2 (2011) ("The search premise of Google at its founding was that it would build an unbiased search engine that consumers would see the most relevant search result first, and that the search results would not be influenced by the web page's commercial relationship with Google." (statement of Hon. Herb Kohl, Chairman, S. Subcomm. on Antitrust, Competition Policy & Consumer Rights)); FRANK PASQUALE, *THE BLACK BOX SOCIETY* 66 (2015) ("We trust our search engines to play straight with us: to show us what's there; to put the best suggestions on top so that we don't have to click through thousands of pages to find them; and to rank by relevance unless they tell us otherwise. But do they?"); accord Daniel Crane, *Search Neutrality and Referral Dominance*, J. COMP. L. & ECON. 1, 9 (2012) (noting the neutrality viewpoint and Google's statements about neutral aspects of its approach to search). This romanticism implicitly assumes technologies are neutral, when they rarely, if ever, are. Drawing on science and technology studies, Julie Cohen argues that we should "reject the assumption that technologies and artifacts have fixed forms and predetermined, neutral trajectories." JULIE COHEN, *CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE* 26–27 (2012). Thus, insofar as search methods in any realm are "technologies and artifacts," claims to neutrality or the level of objectivity to which math and science proofs aspire are suspect.

41. See, e.g., Google Inc., Amendment No. 9 (Form S-1) (Aug. 18, 2004), [hereinafter Google Registration Statement] (stating the search results "are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating").

history also reveals the problem with the view of objective or neutral information sorting and the idea of a perfect or neutral search result.⁴² Recall Vannevar Bush's point that the amount of information is growing and being recorded but "we can hardly consult it," because we have poor tools—akin to the same ones used in the days of square-rigged ships—to find the information we need.⁴³ According to Bush, "[o]ur ineptitude in getting at the record is largely caused by the *artificiality* of systems of indexing and storage."⁴⁴ And as Bush noted, "[t]he human mind does not work that way. It operates by association."⁴⁵ But before we can get to information provision as association, we have to look at indexing in the digital context to see where it worked, what was somewhat objective, and why it no longer suffices.

Despite Bush's point about the limitations of indexing, the practice allowed people to consult the record on the early web, because the number of pages was small. When Yahoo! initially offered its directory, the web was indexed and organized like a thesaurus.⁴⁶ Someone had to decide how to categorize the web and arrange links by hand in a hierarchical order. Yahoo! decided what went where, and websites lived with that decision. Plus, Yahoo! highlighted certain sites when it made its picks of the day—daily recommended sites for users to visit.⁴⁷ Being chosen for that highlighted position would boost traffic and was seen as an accomplishment.⁴⁸ As the web grew quickly, the directory approach could not, however, keep pace.

Search engines entered to find a fast, scalable way to respond to queries. Three search engines, InfoSeek, Lycos, and AltaVista, launched between 1994 and 1995.⁴⁹ AltaVista "fully indexed all of the text on every page of the web—and, even better, results were returned in the blink of an eye."⁵⁰ AltaVista had an index and a superior ability

42. See *supra* notes 5–6 & 40 (discussing view of neutral results in search, news, and shopping).

43. Bush, *supra* note 1, at 42.

44. *Id.* at 43 (emphasis added).

45. *Id.*

46. Cf. Michelle Dalmau et al., *Integrating Thesaurus Relationships into Search and Browse in an Online Photograph Collection*, 23 LIBRARY HI TECH 425, 431 (2005) (noting difficulties in using a thesaurus approach to organizing information and praising Yahoo!'s fourteen top-level term directory structure).

47. Cf. *id.* 450 (noting the collection of photographs being discussed were a Yahoo! Pick of the Day as partial evidence of the success in making the collection findable).

48. *Id.*

49. See, e.g., MACCORMICK, *supra* note 14, at 12.

50. *Id.*

to match searches to web pages.⁵¹ That allowed AltaVista to be the market leader for a few years after its launch.⁵² But there are two keys to successful search services: matching (finding the best hits) and ranking (displaying the hits in the most useful order).⁵³ The modern use of HTML links and metatags started to address Bush's dream of a machine that could use association to aid in consulting and accessing the ever-increasing record of information.⁵⁴ And it also opened the door for the breakthrough for ranking: using a link-analysis algorithm to determine search results. Google's PageRank is the most famous link-analysis algorithm, but most major search companies use the general method in some form.⁵⁵ That said, trying to provide information in any context raises a key point about the nature of information provision in general.⁵⁶

From the information science perspective, providing information involves subjective and objective components.⁵⁷ A look at web search

51. *Id.* at 12, 23 (“The metaword trick did help AltaVista succeed—where others had failed—in finding efficient matches to the entire web. We know this because the metaword trick is described in a 1999 U.S. patent filing by AltaVista, entitled “Constrained Searching of an Index.” However, AltaVista’s superbly crafted matching algorithm was not enough to keep it afloat in the turbulent early days of the search industry.”).

52. *Id.*

53. *Id.* at 11 (“A good search engine will not only pick out the best few hits, but display them in the most useful order—with the most suitable page listed first, then the next most suitable, and so on.”).

54. As with other aspects of the history of information, the ideas of hypertexting and cross-indexing have been around for some time, but Bush is often seen as “the inventor of the modern concept of hypertext.” See Rosemary Simpson et al., *50 Years After “As We May Think”*, 3 ACM INTERACTIONS, Mar. 1996, at 47, 50–51.

55. See EASLEY & KLEINBERG, *supra* note 17, at 412 (noting that although PageRank and an equivalent definition of PageRank may differentiate on the surface, they lead to the same definition).

56. Computer scientists are aware that link analysis only gets one so far in providing search results. See, e.g., MACCORMICK, *supra* note 14, at 36 (“It’s also worth noting that commercial search engines determine their rankings using a lot more than just a link-based ranking algorithm like PageRank.”).

57. Accord Grimmelmann, *supra* note 3, at 915–16 (“[S]earch results are neither entirely ‘objective’ nor entirely ‘subjective.’ The confusion that has surrounded the question for a decade is a result of conflating users’ and search engines’ views of relevance.”). Grimmelmann has offered this analogous point as part of the idea that search is an advisor to the user. See *id.* at 874. He notes that search involves both the subjective part as the input from the user and the objective part as the search providers “expert judgment” used to report back to the user. *Id.* A difficulty arises when one asks what the relationship between user and provider is. Grimmelmann thinks the role is that of “trusted advisor.” *Id.* For him the issue is what the user understands about the information provider. For example, a user of DuckDuckGo, which does not track users as a way to enhance privacy, will be trusted to deliver results that are not personalized and that rely on some other method of sorting. In that case, DuckDuckGo may be expected to use general preferences, link analysis, and other metrics to deliver results in line with what “most users would have preferred to see.” *Id.* (arguing that a provider that “shows tomato.com but knows that most

helps see how that is so. As Sergey Brin and Larry Page, then PhD students, offered in their paper behind Google, “[t]he importance of a Web page is an *inherently subjective matter*, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages.”⁵⁸ The focus of the paper, the initial Google offering, and the idea that captured many people’s imagination was the second idea. That idea is the strategy of leveraging the structure of the early web to rate web pages “objectively and mechanically, effectively measuring the human interest and attention devoted to them.”⁵⁹ The idea was further asserted in Google’s IPO documents.⁶⁰ Of Silicon Valley’s different examples of hubris, this one is the overstated claim that the early link-analysis-based search approach was inherently superior to all other methods, because it was objective or neutral due to the math and science behind it.⁶¹ At most, it was superior as compared to directory or other search approaches at that moment. As in Bush’s era, indexing ran into a wall. The index approach had to address the fact that the web was not static and instead was growing at an incredible rate. Even if matching systems were good, they were hindered by the rate at which they could index. This fact raised an economic problem.

Information services are expensive. In a digital, searchable world, one not only has to index the web but also has to have a backend that returns matched and ranked results to millions, and in some cases billions, of users with different searches at the same time in milliseconds.⁶² Social networks must also handle continual intake of

users would have preferred to see tomato.org, disservices users because it deceives them. The search results are not wrong in an absolute sense, but they are dishonest in the context of [the provider’s] relationship to its users.”). At the same time, one might trust the information provider to try and deliver what the user finds most relevant by using personalization and other methods that have greater subjectivity. I thank Professor Grimmelmann for clarifying his views on what a trusted advisor might be. Any mistakes in this description are mine, not his.

58. LAWRENCE PAGE, SERGEY BRIN, RAJEEV MOTWANI & TERRY WINOGRAD, THE PAGERANK CITATION RANKING: BRINGING ORDER TO THE WEB 1 (1998), <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf> (emphasis added).

59. *Id.*

60. See Google Registration Statement, *supra* note 41, at 32 (stating the search results “are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating”).

61. Drawing on science and technology studies, Julie Cohen argues that we should “reject the assumption that technologies and artifacts have fixed forms and predetermined, neutral trajectories.” COHEN, *supra* note 40, at 26–27. Thus, insofar as search methods in any realm are “technologies and artifacts,” claims to neutrality or the level of objectivity to which math and science proofs aspire are suspect.

62. Danny Sullivan, *Google Still Doing at Least 1 Trillion Searches Per Year*, SEARCH ENGINE LAND (Jan. 16, 2015, 9:00 AM), <http://searchengineland.com/google-1-trillion-searches->

new data from millions, store it, and make it quickly available to other users. In short, any information provider must find a revenue stream. Even Mozilla, operator of the Firefox browser and a nonprofit that has tried to stay away from advertising, has had to start to use advertising to generate revenue.⁶³ Staying within the search industry, one approach, paid inclusion, is when a search engine charges a site to be included in the index.⁶⁴ That method improves the possibility of being ranked but does not guarantee a certain ranking outcome.⁶⁵ As late as 2001, many search engines such as AltaVista and Inktomi engaged in paid inclusion to help pay the bills. According to Danny Sullivan, “Google was the main holdout among the major search engines.”⁶⁶ The difference may have been that Google had figured out how to use an auction to serve ads next to search (an innovation that Overture pioneered, but Google applied quite well), and as such Google did not have to use paid inclusion as part of its funding stream.⁶⁷ Google could argue that its system was superior, in part because it was not taking “payment for [search results] or for inclusion or more frequent updating.”⁶⁸ That is, the approach indexed more of the web more often and did not charge for inclusion in the index that fed what is now known as organic search. But just as the matching approach to search had to evolve to face new realities, so did link analysis and search in general. Thus, a related, mistaken view of many (but perhaps not Brin and Page) is the idea that PageRank, the web, and what allowed link analysis to work—in that sense, search specifically and information provision in general—would stay the same over time.⁶⁹

per-year-212940.

63. See Steven J. Vaughn-Nichols, *Mozilla Quietly Deploys Built-in Firefox Advertising*, ZDNET (Sept. 11, 2015, 6:05 PM), <http://www.zdnet.com/article/mozilla-gets-built-in-firefox-advertising-rolling/>.

64. Danny Sullivan, *Once Deemed Evil, Google Now Embraces “Paid Inclusion,”* MARKETING LAND (May 30, 2012, 9:15 AM), <http://marketingland.com/once-deemed-evil-google-now-embraces-paid-inclusion-13138>.

65. *Id.*

66. *Id.*

67. See GAUTAM SHROFF, *THE INTELLIGENT WEB: SEARCH, SMART ALGORITHMS, AND BIG DATA* 50–51 (2013) (explaining origins of keyword auctions, issues with its initial implementation, and Google’s use of second-price auction to improve the method).

68. See Google Registration Statement, *supra* note 41, at 32 (stating the search results “are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating”).

69. Cf. MACCORMICK, *supra* note 14, at 36 (“Even in their original, published description of Google back in 1998, Google’s cofounders mentioned several other features that contributed to the ranking of search results. As you might expect, the technology has moved on from there: at the time of writing, Google’s own website states that ‘more than 200 signals’ are used in assessing the importance of a page.”).

The problem is that once search became important to how people found things online, game-theoretic issues took over.⁷⁰ People worked to be at the top of the web search results.⁷¹ The authors of early information retrieval targets—scientific papers or newspaper articles—“were not overtly writing their papers or abstracts with these search tools in mind.”⁷² In contrast, web pages are written “with search engines quite explicitly in mind.”⁷³ The methods used for web pages caused one expert to say, “the documents are actively behaving badly.”⁷⁴

Thus the moment when link analysis could reveal the over-revered wisdom of the crowds was soon over.⁷⁵ That ideal works as long as each guess is independent of other guesses.⁷⁶ On the web, once people actively tried to move up in rank, the interaction became a game.⁷⁷ Instead of being able to rely on independent websites signaling quality via links, search providers had to address the rise of search engine optimization, which is now the norm.⁷⁸ As such, “for search engines, the ‘perfect’ ranking function will always be a moving target,” because if the function is static too long, optimizers will game the system to ensure that they are the top of the results.⁷⁹ In short, the search engine is less able to do what it purports to do: help provide what it thinks is the most relevant result for a given user.⁸⁰ This point returns us to the first part of Brin and Page’s points about search, the idea of subjective search results, which they addressed as personalized PageRank or a

70. See EASLEY & KLEINBERG, *supra* note 17, at 414.

71. Cf. MACCORMICK, *supra* note 14, at 36 (discussing web spam and link farms that alter link analysis and the search algorithms’ ability to operate well).

72. See EASLEY & KLEINBERG, *supra* note 17, at 414.

73. *Id.*

74. *Id.*

75. *Id.* at 703–04; accord Gerry Smedinghoff, *The Art, Philosophy, and Science of Data*, CONTINGENCIES, May–June 2007, at 37, 40 (“[T]he crowd is almost always collectively wiser than an individual expert, provided four conditions hold: [1] Diversity: Each person adds private information or bias. [2] Independence: People form their opinions independently. [3] Decentralization: People draw on their own specialized knowledge. [4] Aggregation: A mechanism exists to turn private judgments into a collective decision.”).

76. EASLEY & KLEINBERG, *supra* note 17, at 703–04; accord Smedinghoff, *supra* note 75, at 40 (noting that a Google search does not tap “into the wisdom of any single expert,” instead, the search is “taking a poll of the collective wisdom of websites”).

77. See EASLEY & KLEINBERG, *supra* note 17, at 364–65.

78. *Id.* at 365. Although an ardent critic of Google, Frank Pasquale admits, “[t]he more Google revealed about its ranking algorithms, the easier it was to manipulate them. Thus began the endless cat-and-mouse game of ‘search engine optimization,’” which in turn requires the secrecy he criticizes. PASQUALE, *supra* note 40, at 64–65.

79. See EASLEY & KLEINBERG, *supra* note 17, at 365.

80. See *id.*

personalized search engine.⁸¹

Although today's information practices involve more than simple inputs, from the start, Google had its eye on what it, and others such as Apple, Amazon, Facebook, and Netflix pursue today: using numerous signals (sometimes more than 200) and historical data to address the "*inherently subjective*" or personalized aspect of information retrieval.⁸² Subjectivity here relates to providing results that a specific user finds relevant. The information provider must try to assess what a specific user thinks is relevant.⁸³ The information provider must also make a guess (arguably subjective as it relates to the provider) as to what the user finds relevant.⁸⁴ As Brin and Page said, "search engines could save users a great deal of trouble by efficiently *guessing* a large part of their interests given simple input such as their bookmarks or home page."⁸⁵ This approach has expanded and applies to more than just search.⁸⁶

Social information provision has also followed a path from directory and indexing to a personalized one. For example, social networking on Facebook evolved from a directory of profile pages to the News Feed approach.⁸⁷ In the directory approach, Facebook users posted and other users had to go to another's page to look for updates.⁸⁸ Facebook engineers noticed that some users would look for the page that showed updates to their friends' pages and then go to their friends' pages.⁸⁹ That prompted the creation of News Feed, which gave everyone "a personalized list of news stories throughout the day."⁹⁰ Users would now know when friends changed their pages rather than having to go to their pages to see whether a change had happened.⁹¹ As with other information provision, the amount of information that could be pumped through News Feed is large, so Facebook uses a set of algorithms called Edge Rank and machine learning to determine what goes into a given

81. PAGE, BRIN, MOTWANI & WINOGRAD, *supra* note 58, at 12 ("Such personalized page ranks may have a number of applications, including personal search engines.").

82. *See supra* note 69.

83. *Cf.* SHROFF, *supra* note 67, at 60 (describing a "Reverse Turing Test" where computers try to assess the nature and intent of users to provides ads, web search, social networking, and other services).

84. *Id.*

85. PAGE, BRIN, MOTWANI & WINOGRAD, *supra* note 58, at 12 (emphasis added).

86. *Cf.* SHROFF, *supra* note 83.

87. *See* Victor Luckerson, *Here's How Facebook's New Feed Actually Works*, TIME (July 9, 2015), <http://time.com/3950525/facebook-news-feed-algorithm/>.

88. *Id.*

89. *Id.*

90. *Id.*

91. *Id.*

user's News Feed.⁹²

In general, as users look for posts and news on Facebook, songs on iTunes, tweets on Twitter, films on Netflix, goods on Amazon, local services on Yelp, and so on, each company must offer results that fit a given user's query and interests.⁹³ The need to guess what a given user desires leads us to the idea that individuation—identification of someone as a singular person—can enable individuality.⁹⁴ That is, with deeper understanding of who a given user is, an information provider can provide information connected to and in line with that person's ability to be an individual.⁹⁵

II. THE INFORMATION USER: HOW INDIVIDUATION CAN FEED INDIVIDUALITY

Frank Webster captures the interaction between individuation and individuality: “[I]f we as a society are going to respect and support the individuality of members, then a requisite may be that we know a great deal about them.”⁹⁶ Webster's point was about state surveillance and provision of entitlements that allow someone the “capacity to be true to themselves” and exercise “genuine choice.”⁹⁷ Peter Swire's work on

92. *Id.*

93. The evolution of Facebook reflects a shift in approaches to social networks and information provision in that arena. How information providers choose to organize results has, nonetheless, been the subject of lawsuits claiming defamation, extortion, and deceptive business practices in some cases. *See, e.g.,* *Levitt v. Yelp, Inc.*, 765 F.3d 1123 (2014); *Reit v. Yelp, Inc.*, 907 N.Y.S.2d 411 (2010). So far courts have sided with information providers' ability to arrange results as they “see fit” even if such choices pose a “threat of economic harm” to given company. *Levitt*, 765 F.3d at 1134; *cf. Reit*, 907 N.Y.S.2d at 718 (dismissing complaint where damages claimed were lost business because of Yelp's choice in ordering reviews). And arguments that an information provider may have ordered results to aid the selling of advertisements have been taken as part of business practices that do not harm consumers. *See Levitt*, F.3d at 1134 (“As Yelp has the right to charge for legitimate advertising services, the threat of economic harm that Yelp leveraged is, at most, hard bargaining.”); *Reit*, 907 N.Y.S.2d at 718 (“Reit's allegation that Yelp deletes postings for the purpose of selling advertising, if true, is business conduct, not consumer-oriented conduct.”).

94. WEBSTER, *supra* note 15, at 55; *accord* Grimmelmann, *supra* note 3, at 899 (“It is precisely because people have wildly diverging needs, capabilities, values, preferences, worldviews, and life experiences that the individuation of search matters. . . . The development of personalized and social search is not just a means towards ‘better’ results, it is also a way of accommodating diversity of user interests. One man's noise is another man's signal; delegating to users the decision of what to search for lets them make different decisions.”).

95. Grimmelmann, *supra* note 3, at 899.

96. WEBSTER, *supra* note 15, at 55.

97. *Id.* at 175, 208; *accord* Julie E. Cohen, *What Is Privacy for*, 126 HARV. L. REV. 1904, 1917 (2013) [hereinafter Cohen, *What Is Privacy for*] (“Governments require some kinds of knowledge about people to govern effectively.”). The privacy implications of data collection and individuation are beyond the scope of this Essay, but they merit mention. Webster's analysis held that the individuation was not fully tied to a given person, but that is no longer the case.

fair lending law's implications for marketing illustrates the paradox.⁹⁸ Targeted marketing can involve gathering large amounts of consumer data and might be used to deny certain groups access to vital services such as loans. Yet, targeted marketing, that is, individuation used to know that one is reaching a minority, is a remedy for denial-of-lending opportunities, because society can order a lender to offer services to groups otherwise left out.⁹⁹ In Webster's terms, if we take lending to be one of those practices that "respect[s] and support[s]" individuality, we may need to know to whom we are lending and even be active in reaching some groups. Webster connected the issue to what he called "corporate capitalism," and the general structure fits our question.¹⁰⁰

If one demands that information filters and recommenders somehow deliver what each one of millions, or in some cases billions, of users

WEBSTER, *supra* note 15, at 57. Several scholars have addressed how highly specific individuation has potential negative implications for capacity and choice as a matter of privacy. Cohen, *What Is Privacy for*, *supra*, at 1912 (arguing that surveillance diminishes the capacity for "democratic self-governance" because individuals no longer have the space to develop their version of citizenship rather than one dictated or shaped by the state); Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1391–92 (2000) [Cohen, *Examined Lives*]; Deven R. Desai, *Constitutional Limits on Surveillance: Associational Freedom in the Age of Data Hoarding*, 90 NOTRE DAME L. REV. 579, 619–25 (2014). See generally COHEN, *supra* note 40 (examining the problems of data, freedom from surveillance, and self-governance). Nonetheless, there is a paradox because increased surveillance for state control or capitalist goals can also "increase choices for people." WEBSTER, *supra* note 15, at 156–57 (referencing Anthony Giddens's books *The Consequences of Modernity* and *Modernity and Self-Identity: Self and Society in the Late Modern Age*).

98. PETER SWIRE, LESSONS FROM FAIR LENDING LAW FOR FAIR MARKETING AND BIG DATA (2014), https://www.ftc.gov/system/files/documents/public_comments/2014/09/00042-92638.pdf.

99. *Id.* at 6–7.

100. WEBSTER, *supra* note 15, at 73. The private side of surveillance is real. See, e.g., Spiros Simitis, *Reviewing Privacy in an Information Society*, 135 U. PA. L. REV. 707, 773 (1987) ("The broad availability of personal data and [] elaborate matching procedures [mean] individual activities can be accurately reconstructed through automated processing."). This issue has prompted Neil Richards to call for protecting "the ability to develop ideas and beliefs away from the unwanted gaze or interference of others." Neil Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 389 (2008). Julie Cohen's work on privacy and play links the problems of data, freedom from surveillance, and self-governance. See generally, COHEN, *supra* note 40. We want people to have the "capacity for critical independence of thought and judgment," "self-actualization and reason," and "cosmopolitanism," because these capacities allow people to be full citizens of our society who can "identify" and "pursue" their personal and political self-fulfillment. Cohen, *What Is Privacy for*, *supra* note 97, at 1911. But those capacities need room to develop, and in that sense we play. *Id.* We explore ideas and "boundaries" of roles and social rules. *Id.* Even though personalization and individuation can enable better search and allow us to make better use of the record, the need for Cohen's room to play and what Richards calls intellectual privacy is important as well. There is thus a distinction between open and complete data on a user (and how much individuation is required to enable increased individuality or capacity) and the need for open networks that allow users to explore all that is available to them. See *infra* CONCLUSION.

think is relevant, personalized search and filters become necessary.¹⁰¹ Some argue, however, that because search and recommendation systems have become personalized, users have lost control over what they see in ways that are harmful to individuality because of the so-called “filter bubble.”¹⁰² The claim is that the current ways in which information is organized—either through a user employing filter tools, a provider filtering, or both—means that we enable a sort of deaf, dumb, and blind behavior; that filters reinforce predilections and harm public debate,¹⁰³ because of “increased insularity and reinforced prejudice.”¹⁰⁴ Regarding news, a racist or liberal or libertarian or anyone with a particular leaning would be expected to use filters to ensure that he or she sees only news that aligns with and reinforces his or her current political position. A related idea is that an information provider can use filters to serve results that benefit the provider, not the user.¹⁰⁵ There are several problems with these views of personalization, not the least of which is the paradox that exposure to more views can lead to less, not more, political engagement. To understand how that is so, we must address some assumptions behind the filter bubble view and look at evidence of whether, and what happens when, people are exposed to views differing from their own.

A. *The Quest for Relevancy and Maybe Serendipity*

Relevancy in the information context is quite personal. Relevancy is personal, because each user determines which results are relevant, and because the more personal information a user shares with an information provider, the more likely a result will be relevant to the seeker but not necessarily for other seekers. Consider, for example, a search by category and type such as choosing a restaurant. If you had a craving for Chinese food, you might ask a friend about a Chinese restaurant. If she responded with the names of some Turkish restaurants, you might think that she was not listening and that she was unhelpful. Instead, you might turn to recommendation systems (e.g.,

101. Cf. PASQUALE, *supra* note 40, at 78–79 (expressing concern that what he sees in a search is not the same as what someone else sees for the same search).

102. *Id.* at 79.

103. The claim is that the ability to have personalized or tailored information fosters “fragmentation, polarization, and the destruction of the possibility of common discourse in the public sphere.” Yochai Benkler et al., *Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate*, BERKMAN CTR. FOR INTERNET & SOC’Y 6 (2013), (citing SUNSTEIN, *supra* note 36).

104. PASQUALE, *supra* note 40, at 79.

105. This section addresses this concern from the user point of view. The discussion below addresses it from the information provider and learning view. See Part III *infra*.

Yelp), searches (e.g., Bing or Google), social networks (e.g., Facebook or Pinterest), and other tools may help you narrow the Chinese restaurants in your area, see what others thought about them, and even discover what your friend thought (if she posted reviews or pinned pages about Chinese restaurants) regardless of her tendency to be nonresponsive. If results gave you names you did not know, you would be surprised. We might say that the previously unknown option was serendipitous. But serendipity is a seductive, overstated idea.

Serendipity works because of relevancy. An assumption in filter bubble and echo chamber critiques is that it is better to be surprised and have serendipitous, “involuntary” encounters with new ideas.¹⁰⁶ Truly random encounters are not, however, always pleasant, let alone useful. Who would not want a pleasant surprise from which one learned something new? Yet that idea already softens serendipity to include learning something new rather than pure randomness. To claim that information should offer serendipity is to claim to know what bit of information may be of interest while also decreasing the sameness of results. The problem with this approach is that not all users are the same, *and* even the same user will have different desires depending on her specific context when seeking information.¹⁰⁷ It also hides the type of serendipity at stake. The serendipity argument is really a call for better exposure to relevant, but unknown information.

Returning to our search for Chinese food, when you believe your friend tends to make good suggestions, you may go to her before consulting online resources, knowing that a request for Chinese food will yield a list of restaurants from Indonesian to Vietnamese. You may recognize that she knows you have visited Turkey, love Turkish food, but rarely find Turkish restaurants you like. If so, a suggestion for Turkish food is not random. The more she knows you and what you are looking for, and the more you have experienced success with your interactions with her, the more otherwise odd suggestions will not be odd. You may be happy with the offer based on experience and shared tastes at a more abstract level. She may be exactly the person to whom you go, because her informed offerings are relevant to you, though few others would find them to be so. Although it appears she has forced you to consider an option that you did not want, that is not the case; she gave you an informed, difficult to discern option that you had reason to

106. Cf. SUNSTEIN, *supra* note 37, at 15 (arguing that traditional news outlets enabled exposure to opposing views “quite involuntarily”).

107. Accord Hunter, *supra* note 4, at 627–34 (noting different ways that mistakes and changes in individual tastes and interests defeat desires for “perfect filtering”).

consider. In contrast someone who repeatedly said you must try vegan restaurants or gluten-free pizza or spicy food when you have no interest or need for those options would not be offering options about which you cared, despite the possibility that they thought the options were good from their view. In other words, defining relevancy is not a simple task.

Thus, an open question is exactly who should decide to what someone should be exposed.¹⁰⁸ Historically, we have regulated over-the-air transmissions such as radio and television regarding what has to be shown or carried based on the idea that the airwaves are a scarce resource with room for only a few to broadcast.¹⁰⁹ But as Jonathan Nuechterlein and Philip Weiser point out, once scarcity was not the case, “the government’s justification for playing its traditional spectrum-management role . . . largely evaporate[d].”¹¹⁰ Today, with digital cable, satellite, and Internet media offerings, broadcast-era scarcity is not at hand. Nonetheless, we turn to editors, station programmers, bloggers, online news, search services, social networks, and more to tell us to what we should pay attention and what is going on in the world. If we know that a service seems fair or presents the news about which we care, we are likely to return. But to say that those editors, station programmers, bloggers, online news, search services, social networks, and more must take space to offer low-interest stories or opposite views of the core audience is still to require someone to decide what those other offerings should be.¹¹¹

Although some seem quite confident to say they know what is good information or what contrary information should be presented to people, this position ignores the fact of filtering and that choice lies at the heart of such decisions. The trade is not for the open, neutral provision of

108. Grimmelmann captures the problem with the forced viewing aspect of filter bubble arguments well: “[T]he consequences of forcing search users to look at results they didn’t ask for and don’t want to see are dreadful. It turns users into Alex from *A Clockwork Orange*, forcibly subjected to high culture and unpleasant truths.” See Grimmelmann, *supra* note 3, at 908.

109. See, e.g., JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, *DIGITAL CROSSROADS: AMERICAN TELECOMMUNICATIONS POLICY IN THE INTERNET AGE* 229–30 (2005). The “assumption that [the spectrum] is ‘scarce’—that there is less of it than an unregulated public could use without causing serious interference problems” was the logic behind the Supreme Court’s view that the FCC could “serve as the traffic cop of the airwaves, . . . [and] ‘determin[e] the composition of th[e] traffic’ permitted on particular slices of the spectrum.” *Id.* (citing to the Court’s decisions in *NBC v. United States*, 319 U.S. 190 (1944), and *Red Lion Broadcasting Co. v. FCC*, 395 U.S. 367 (1969)).

110. *Id.*

111. Cf. Thomas W. Hazlett, Sarah Oh & Drew Clark, *The Overly Active Corpse of Red Lion*, 9 NW. J. TECH. & INTELL. PROP. 51, 81–84, 92–93 (2010) (describing how both political parties used *Red Lion*’s logic and the Fairness Doctrine to chill speech and allow government choices over content).

information. It is a substitution of one set of views about what and how to filter for another. The problem of each person's individuality clashes here. For everyone may have a different idea about what news they should see when looking for information.

Even if your goal were somewhat general, if the information provided had little to do with your inquiry or interest, but you clicked on those materials, you would be frustrated rather than helped.¹¹² More concretely, forced or increased potential for serendipity in all cases when we seek something specific often leads to information fatigue; it leads back to calls for filtering and help getting around the random information thrown at us.

That wonderful things may come of serendipity is a truism. That many information users *and providers* may want, need, or prefer some direction is apparent as well. What may be surprising is that online behaviors are not as blinkered as some would argue. In addition, exposure to opposing views may have little effect in changing the current polarization of politics that many fear filtering either creates or furthers.¹¹³

B. User Tensions: Isolation, Connection, and Action

Despite continued claims that new media creates “inward looking and extreme” clusters of like-minded people, recent empirical work indicates that evidence of such outcomes is lacking.¹¹⁴ It is well known that like-minded, politically engaged people cluster and engage in “selective exposure” in general and on the Internet.¹¹⁵ Yet, rather than

112. Cf. Hunter, *supra* note 4, at 630–31 (noting that recommendations assuming one likes Hugh Grant movies when in fact one likes a certain writer would be incorrect and not useful).

113. See, e.g., Henry Farrell, *The Consequences of the Internet for Politics*, 15 ANN. REV. POL. SCI. 35, 40 (2012) (“American politics is becoming increasingly polarized between left and right. The causes for this polarization are the topic of vigorous debate.”); Greg Lukianoff & Jonathan Haidt, *The Coddling of the American Mind*, ATLANTIC (Sept. 2015), <http://www.theatlantic.com/magazine/archive/2015/09/the-coddling-of-the-american-mind/399356/> (describing increasing polarization of politics from the 1970s when “mutual dislike” between Republicans and Democrats was “surprisingly mild” to the 2000s when “affective partisan polarization” grew and now poses “a very serious problem for any democracy”).

114. See Farrell, *supra* note 113, at 41 (“There is emphatic evidence that politically engaged content producers cluster with others who share their political orientations. However, it says little about the consequences of clustering. Sunstein claims both that the Internet encourages like-minded people to find each other and that when they do, they will reinforce each other’s beliefs so that they become inward looking and extreme. Although the evidence clearly supports the first claim, it is insufficient to support the second.”) (citing sources); *but see* Stroud & Muddiman, *supra* note 4 at 617 (“numerous research studies have found that people prefer information matching their political proclivities, a behavior known as selective exposure. Several studies focusing on online behavior illustrate that the same phenomenon occurs online.”).

115. See Farrell, *supra* note 113, at 41.

people being fully purposeful and seeking and filtering to the point of nonexposure, studies of blog readership, social media, and online news consumption show that people are not isolated in their exposure to differing views.¹¹⁶ A study of political blogs found “no support” for the idea that the Internet would make for increasing isolation and fragmentation.¹¹⁷ A study of Twitter found that despite clustering, a politically minded user “would be exposed to content from users on both sides of the political spectrum.”¹¹⁸ Even if one discounts bloggers and Twitter users as a special, engaged subset, other work shows that readers of nonpolitical forums encounter political views different than their own in those forums.¹¹⁹

In addition, newsreaders are also less isolated than claimed. According to one study, “most online news consumption is concentrated in a small number of relatively centrist sites.”¹²⁰ Although “the ‘long tail’ of political blogs, news aggregators, and activist sites . . . are often ideologically extreme . . . they account for a very small share of online consumption.”¹²¹ Furthermore, the study also found that “a significant share of consumers get news from multiple outlets.”¹²² One might think that this description is for those who frequent centrist sites. The multiple outlet behavior, however, was

especially true for visitors to small sites such as blogs and aggregators. Visitors of extreme conservative sites such as rushlimbaugh.com and glennbeck.com are more likely than a typical online newsreader to have visited nytimes.com. Visitors of extreme liberal sites such as thinkprogress.org and moveon.org are more likely than a typical online newsreader to have visited foxnews.com.¹²³

It appears that rather than being sheltered, isolated readers, those who seek political news are reading opposing views.¹²⁴

116. *Id.* As discussed below, some find the issue to be how much an information provider such as Facebook allows a user to alter the amount of news with which they disagree rather than the fact that like-minded people tend to read and share news with which they agree. See *infra* note 216 and surrounding discussion.

117. See Eszter Hargittai et al., *Cross-ideological Discussions Among Conservative and Liberal Bloggers*, 134 PUB. CHOICE 67, 85 (2008).

118. M.D. Conover et al., *Political Polarization on Twitter*, 5 INT’L AAAI CONF. ON WEB & SOC. MEDIA 89, 94 (2011); accord Farrell, *supra* note 113, at 41.

119. Farrell, *supra* note 113, at 41.

120. Matthew Gentzkow & Jesse M. Shapiro, *Ideological Segregation Online and Offline*, 126 Q.J. ECON. 1799, 1801 (2011).

121. *Id.* at 1802.

122. *Id.*

123. *Id.*

124. Here we can see connections between filter criticisms regarding news and older, broader claims that consumer capitalism isolated people into the home. The older concerns were that

That people are reading, not just seeing, other news does not tell us what they are doing with that news and whether their views are changing. As the authors of the news study caution, “none of the evidence here speaks to the way people translate the content they encounter into beliefs. People with different ideologies see similar content, but both Bayesian and non-Bayesian mechanisms may lead people with divergent political views to interpret the same information differently.”¹²⁵ Whether one is open to challenging ideas can turn on a situation, that is, whether there is a “perceived threat” akin to “tackling troubling situations or making critical decisions that are either personal, such as those regarding security, health, finance, or societal, such as crime, terrorism, economic crisis.”¹²⁶

Openness to new information can also turn on someone’s degree of involvement with a topic.¹²⁷ One small study found that “even when opposing views were presented side-by-side,” an information seeker “under perceived relevant threat” would have a “more pronounced selective exposure to attitude consistent information.”¹²⁸ Thus, change in attitude was less, because they had received “overall” less “attitude challenging information.”¹²⁹ We might think that someone “tackling troubling situations or making [a] critical decision” should be discounted because of the motivation behind their search. Yet, those who were highly involved with a topic fared little better regarding a change of attitude. Although that group sought “relatively balanced exposure to attitude consistent and inconsistent information,” they still showed “more preferential evaluation of attitude consistent information over attitude inconsistent ones.”¹³⁰ Indeed high involvement “largely increases the resistance to attitude change.”¹³¹ These points leave open the question of what action, if any, follows from the evidence that

there was only “garbage information” and people were allowed to consume only “rubbish”; yet, as Webster notes, there were several objections to that view. See WEBSTER, *supra* note 15, at 98–99. First, there was an increase in a range of quality of overall programming and access to and readership of classics had improved thanks to low-cast paperbacks. *Id.* at 98. And even in the days of mass-market reading as a main form of entertainment, the average reader was not reading “Cobbett or Hardy” but “penny dreadful, sensationalized trivia about murder, rape, drink, and fallen women.” *Id.*

125. Gentzkow & Shapiro, *supra* note 120, at 1801.

126. Q. Vera Liao & Wai-Tat Fu, *Beyond the Filter Bubble: Interactive Effects of Perceived Threat and Topic Involvement on Selective Exposure to Information*, 2013 PROC. SIGCHI CONF. ON HUM. FACTORS COMPUTING SYS. 2359, 2359.

127. *Id.* at 2360.

128. *Id.* at 2366.

129. *Id.*

130. *Id.*

131. *Id.*

people are indeed seeing opposing views.

If those who fear that the Internet fosters extremism and isolation ask not only that we all have greater exposure to more ideas but also take action, it turns out that the two goals conflict.¹³² As Henry Farrell explains, the Internet presents a “troubling political dilemma.”¹³³ Someone having connections with those who have differing views may be inclined to *less* political participation.¹³⁴ Although one may support the sort of crosscutting connections (and assumed exposure to opposing views) that is found on the Internet, the clustering or homophily¹³⁵ effect on the Internet—the practice of finding like-minded people that concerns some critics—is precisely the sort of connection and support that an individual and group need for collective action.¹³⁶ This last point leads to the idea of democratic flatness.

The Internet holds open the *possibilities* of openness, populist action, and enhanced democracy, but those possibilities require work to be realized. To think that the creation of the Internet meant “an inevitable, irresistible revolution was just around the corner, one that would flatten society, unseat the elites, and usher in a new kind of free wheeling global utopia”¹³⁷ or would “entirely redemocratize society” is an oddly, and perhaps naive, technological-exceptionalist view of the Internet, if not all media.¹³⁸

A related version of this view is that there are centralizers such as “AT&T, Hollywood, and Apple” to be feared and fought and decentralizers who favor openness and “aspire to nothing less than social transformation.”¹³⁹ The decentralizers claim that their world will be one where “most goods and services are free or practically free, thereby liberating the individual to pursue self-expression and self-actualization as an activity of primary importance.”¹⁴⁰ From this perspective, once an information provider is large enough in one realm,

132. See Stroud & Muddiman, *supra* note 4 at, 620–21 (“[T]he use of counter-attitudinal media is related to lower levels of political participation and less polarized attitudes and that likeminded partisan media use predicts higher levels of participation and polarization.”)

133. Farrell, *supra* note 113, at 42.

134. *Id.*

135. Hargittai et al., *supra* note 117, at 69–70 (discussing the history of homophily).

136. Farrell, *supra* note 113, at 39.

137. PARISER, *supra* note 37, at 3.

138. *Id.* at 5; see also Stroud & Muddiman, *supra* note 4, at 608–09 (discussing how expansion of news sources dropped engagement with specific items (e.g., Presidential debates), but that incidental exposure to news persists regardless of the medium).

139. TIM WU, THE MASTER SWITCH: THE RISE AND FALL OF INFORMATION EMPIRES 296 (2010).

140. *Id.*

the provider must be constrained to stay within that realm.¹⁴¹ For example, Tim Wu has argued that Google is at its best when it stays specialized as a firm that “harvests the best of the Internet, organizing the worldwide chaos in a useful way, and asks its users to navigate this order via their own connections; by relying on the sweat of others for content and carriage, Google can focus on its central mission: search.”¹⁴² Yet, Wu acknowledges that a company that is not “vertical[ly] integrated” and rests on open architecture and connecting others’ work is “underintegrated,” “undefended,” and “vulnerable,” “rather like a medieval city without a wall.”¹⁴³ Wu was concerned with the way telecommunications companies such as AT&T could clamp off Internet access to Google.¹⁴⁴ That position, however, misses that an information provider that is underintegrated, undefended, and vulnerable also faces threats from other information providers. In addition, to remain viable, maintain its user base, and grow, any information provision company must continue to understand users’ interests, personalization, and new modes of information seeking and provision.

There is thus a tension in what may be desired and the outcomes that flow from proscribing what areas an information provider may enter. On the one hand, Wu recognizes that Google had to evolve to combat moves that shut it off from Internet users, and so Google had to launch Android to counter Apple’s moves to create a walled-in iWorld.¹⁴⁵ On the other hand, he rejects the idea that Google should be able to compete on local search.¹⁴⁶ The claim is that Google must not be allowed to take steps to vertical integration and reduce its vulnerability, but instead it must stay a “tool” or “venue of access” to would-be competitors.¹⁴⁷ The logic is similar to the days of the now-defunct *ultra vires* doctrine, where a company declared its business purpose and was not allowed to pursue anything other than that initial business plan.¹⁴⁸ When we think of Wu’s ideas as part of filter bubbles, control, and

141. *Id.* at 283.

142. *Id.*

143. *Id.* at 281–83.

144. *Id.* at 284.

145. *Id.* at 293–94.

146. See Michael Luca et al., *Does Google Content Degrade Google Search? Experimental Evidence* 33 (Harv. Bus. Sch., Working Paper No. 16-035, 2015), http://www.hbs.edu/faculty/Publication%20Files/16-035_53c43fb4-86e8-49e3-a419-db2ab8350694.pdf.

147. Wu, *supra* note 139, at 304.

148. See, e.g., Deven R. Desai, *The Chicago School Trap In Trademark Law: On the Co-Evolution of Corporate, Anti-Trust, and Trademark Law*, 37 CARDOZO L. REV. (forthcoming 2015).

information provision in general, Facebook would not be allowed to provide search, news, and video services, Google could not engage in social networking, Amazon would not be able to expand its offerings, Apple would not be allowed to offer browsers and search, and so on. Competition amongst each would diminish; and if the state mandated such separation, each would be a modern, state-sanctioned monopoly like utilities of old. Either way, these walls would seem to foster filter bubbles and also hinder competition and innovation.

In both the filter bubble and competition criticisms, an underlying claim is that decentralized power via today's technology gives individuals "more power than at any time in the past century," but changes in technology and centralizing forces mean an individual's ability to "hold on to [that power]" is threatened, if not swallowed.¹⁴⁹ Yet, each time a new medium emerges, it presents opportunities for more democratic action and for forces to take over that action.¹⁵⁰ In their best light, concerns about filters and competition rest on caution about this tension in media.

Nonetheless, the claim that filters today are different than the past relies on an underlying misconception that the ability to choose has been given up. Indeed, the evidence of greater exposure to differing views belies claims that someone "trains" an information provider about what "to show us" and that in turn the provider "gradually conditions us [what] to expect [next]" so that we no longer see anything other than that which pleases us.¹⁵¹ In addition, from a practical, information provider point of view, such a strategy would be unwise and counterproductive, and would undermine the provider's ability to succeed.

III. THE INFORMATION PROVIDER: THE NEED FOR EXPLORATION AND EXPLOITATION

Increasing information tends to defeat perfect filtering.¹⁵² The potential for avoiding new information and ideas diminishes as the amount of information and the number of people online increases. Information services seek to provide the best answers for millions of people across a huge range of possible answers at almost every moment of the day. Even in a relatively limited space such as a film library,

149. WU, *supra* note 139, at 298.

150. For a history of media and its relation to politics, see generally PAUL STARR, *THE CREATION OF THE MEDIA: POLITICAL ORIGINS OF MODERN COMMUNICATIONS* (2005).

151. PASQUALE, *supra* note 40, at 78; see also SUNSTEIN, *supra* note 37, at 8, 15–16.

152. *Accord* Hunter, *supra* note 4, at 627–34.

recommending films that match a given person's taste and desires in the moment is quite difficult.¹⁵³ When one turns to spaces such as dining options in a city, sellers of goods, digital libraries, social networks, blogospheres, and of course, the web, the problem of presenting the information a user wants grows larger. In other words, the information world is not as stable as those with filter concerns might believe.¹⁵⁴ Serendipitous information encounter may be baked into information retrieval, because information retrieval is imperfect even as it strives to connect with users. The interaction between exploration and exploitation helps understand why.¹⁵⁵

A. Learning Through Exploration and Exploitation

An information provider is an organization and has a need for organizational learning. As James March has explained, an information provider must use both exploration “of new possibilities,” and exploitation “of old certainties,” to succeed as an organization and to meet the needs of the diverse array of information users.¹⁵⁶ Just as organizations move between exploration and exploitation, people want something new but not completely random and so too participate in exploration and exploitation.¹⁵⁷ An information provider thus confronts a moving target, the user, and must also engage in exploration and exploitation as matter of organizational learning. Furthermore, information science explicitly incorporates both practices as a way to improve learning.¹⁵⁸

153. See Tien T. Nguyen et al., *Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity*, 23 PROC. INT'L CONF. ON WORLD WIDE WEB 677, 685 (2014) (noting that as film viewing diversification increased, user satisfaction of films increased as well).

154. See *id.* (explaining that even within a closed film recommendation system there was “evidence for two forms of narrowing when analyzing all users—the items recommended by the system and the items rated by users both became slightly narrower (less diverse) over time. However, the results for *all* users obscure the most interesting part of the story. The narrowing effect actually was mitigated for users who appeared to ‘follow’ the recommender (operationalized as having rated movies that appear in their top-*n* recommendation lists); in other words, *taking recommendations lessened the risk of a filter bubble.*”).

155. Cf. DOMINGOS, *supra* note 2, at 129 (calling the exploration-exploitation problem “[o]ne of the most important problems in machine learning—and life”).

156. See James G. March, *Exploration and Exploitation in Organizational Learning*, 2 ORG. SCI. 71, 71 (1989); accord David Lazer & Allan Friedman, *The Network Structure of Exploration and Exploitation*, 52 ADMIN. SCI. Q., 667, 671 (2007) (“[E]xploration generally involves attempts to introduce new information, while exploitation leverages existing knowledge for some productive end.”).

157. Cf. Grimmelman, *supra* note 3, at 908 (“There is a deep and abiding human taste for novelty, for serendipity, for the unfamiliar. Search engines are particularly useful in helping people explore new interests quickly and easily.”)

158. See, e.g., DOMINGOS, *supra* note 2, at 129 (connecting exploration-exploitation ideas to

Although exploration and exploitation can be understood as separate activities, they feed on each other. After initial exploration to learn how to proceed, an information provider wants to apply or exploit that knowledge to deliver user results; but to understand whether the application is accurate, misses something, and can be improved, a provider has to explore other options over time.¹⁵⁹ Exploration “of new possibilities,” uses data to learn at a high level and to discern patterns.¹⁶⁰ Demographic analysis for advertising and suggestions systems—such as Pandora for music, Facebook for news and ads, Pinterest for web page and ideas sharing, and Google for web search—are examples of exploration outcomes. This type of exploration can yield the type of serendipity that makes sense.¹⁶¹ It may also lead to exploitation “of old certainties,” which in this example is the focused or targeted use of data such as for personalized news, social network feeds, search results, and advertising.¹⁶² The question may seem to be which option to pursue, but the point is that an organization must pursue both options.

Computer scientist Pedro Domingos explains the dynamic and general dilemma here. At first, one explores thoroughly and “once you know the territory, it’s best to concentrate on exploiting it.”¹⁶³ But staying with that solution may mean one has settled into a “local optimum” that misses better options. As he asks, “If you’ve found something that works, should you just keep doing it? Or is it better to try new things, knowing it could be a waste of time but also might lead to a better solution?”¹⁶⁴ Put differently, one might think learning occurs only in exploration, but that is not quite correct. Recent work on exploration and exploitation has debated whether learning occurs in only exploration and not exploitation, and it appears that each option “includes at least some learning.”¹⁶⁵ The question is the “amount of

machine learning).

159. Time and the nature of the network within which information moves are important to this process. See generally Lazer & Friedman, *supra* note 156, at 689–90. It appears that the longest time horizon should have inefficient communication, because that inefficiency allows for greater learning and exploration. *Id.* Given, however, the demands for quick outcomes, highly connected networks that converge on a solution, but are less high performing, are the norm. *Id.*

160. See March, *supra* note 156.

161. Cf. Grimmelman, *supra* note 3, at 908.

162. See March, *supra* note 156.

163. See, e.g., DOMINGOS, *supra* note 2, at 221 (2015) (connecting exploration-exploitation ideas to machine learning).

164. *Id.*

165. Anil K. Gupta et al., *The Interplay Between Exploration and Exploitation*, 49 ACAD. MGMT. J. 693, 694 (2006).

learning rather than the presence or absence of learning.”¹⁶⁶ As a matter of organizational strategy and information provision, overcommitting to either exploration or exploitation can undermine the benefits of the other.¹⁶⁷ Too much exploration results in high-cost learning “without gaining many of its benefits.”¹⁶⁸ And for an information provider too much exploration risks upsetting users and perhaps pushing them away. Too much exploitation will see only what is known and miss new patterns and problems.

Any company—such as Amazon, Apple, Bing, Facebook, Google, Pinterest, Netflix, Spotify, Twitter, or Yelp—that sorts information and offers results for millions of users, must balance between offering results that are consistent and perhaps boring (feeding short-run outcomes) and offering results that are a bit different and not clearly ones a user has seen before (feeding long-run outcomes). Thus, work building on March’s scholarship found “a single core insight: the more efficient the network at disseminating information . . . the better the system performs in the long run and the worse in the short run.”¹⁶⁹ A company needs to make sure that there is enough information diffusion to allow others to see what the current, known best results are and leads to copying of the current “effective strategies.”¹⁷⁰ That is the short run outcome. At the same time that effect can lead to a lack of diversity; but an information provider must not kill off diversity, because diversity “is also positively related to performance.”¹⁷¹ More specifically, one can take information provision as a series of exploration-exploitation games, “[e]ach time you play, you have to choose between repeating the best move you’ve found so far, which gives you the best payoff, or trying other moves, which gather information that may lead to even better payoffs.”¹⁷² Thus, in some cases, a repeated search for the same term will generate new results, because that behavior is a signal that the initial results, although reflecting what others found relevant, were not

166. *Id.*

167. *See* March, *supra* note 156.

168. *See id.*

169. Lazer & Friedman, *supra* note 156, at 686.

170. *Cf. id.* (“Our analyses thus capture both positive and negative pathways through which the network configuration affects performance. An efficient network positively affects information diffusion, which facilitates the spread of effective strategies, but negatively affects information diversity, which is also positively related to performance.”).

171. *Id.* Even though there is “near consensus on the need for balance” between “exploration and exploitation,” the nature of that reaching that balance is not settled. *See* Gupta et al., *supra* note 165, at 697. The leading ideas are ambidexterity where both activities occur within units in an organization, and punctuated equilibrium, where there are “long periods of exploitation and short bursts of exploration.” *Id.*

172. *See, e.g.,* DOMINGOS, *supra* note 2, at 130.

what the user found relevant. The information provider is in that sense using the interaction to A/B test options and improve both results for that user and its overall system of information provision.¹⁷³ Whether a given person will find those results palatable will vary based on context and the aim of the inquiry. Some examples help illustrate this point.

Let's begin by considering sorting within a specific area such as music. I was listening to a jazz station on Rhapsody, an online music streaming service. I began by choosing one of several jazz stations by seeing what artists Rhapsody listed as representative of the jazz genre offered for each station. Then I filtered further. I set the variety slider, which went from low to high, to medium. While listening, the song *Howling for Judy* by Jeremy Steig came on. I had never heard the song before, but recognized the flute riff. It was used in the Beastie Boys song, *Sure Shot*. I liked Steig's song enough that I clicked on the song to see the album and read about the artist. I then searched for the specific album online and purchased it.

Filters led to an encounter that was in fact highly unlikely in the unfiltered world. The filters connected some dots and generated a more salient serendipity. Yet in my experience, when the variety slider is set to a high level of variety, songs are not at all to my liking and don't fit my view of the subgenre I chose. That makes me think I should change the setting and maybe try other music recommendation services. We can extend the recommendation problem too. If I were listening to the Beastie Boys, I would welcome a service suggesting related tunes, not just from the rap/hip-hop genre but the other genres influencing the song. Yet someone who only wanted to listen to rap/hip-hop might find those suggestions irritating or useless. From an exploration-exploitation view, if one is a genre purist or one wants a broader range of recommendations, the filter setting helps a user signal how much exploring or exploiting information she desires and also aids her exploration and exploitation needs. Nonetheless, if the suggestions were truly random, the so-called serendipity would likely be ignored, because rather than being salient, the suggestions would appear to be irrelevant and unrelated. Thus, suggestions that seem to have no relation to a song or genre of interest raise the question of whether the suggestion system is poor and possibly rigged to push songs in which I have no interest but that have a higher margin for the company.¹⁷⁴ If

173. See *id.* at 227 (describing A/B testing and noting use by Amazon and Google as a core part of their services).

174. Cf. PAGE, BRIN, MOTWANI & WINOGRAD, *supra* note 58, at 12 (noting possibility that commercial interests can cause manipulation of information provision results).

so-called “good” suggestions, as defined by me, are my criteria, Rhapsody’s failures push me to other services. For any recommendation or information provision service, not predicting millions of users’ desires at every moment (exploitation) can be deadly. But not suggesting things that come close to, even if not exactly matching, a given user’s idea of good (exploration) can be deadly as well.

A general search presents a greater challenge for exploration and exploitation. Suppose someone searched the web for Darwin. Take a moment. What result do you think should appear? What information should be retrieved? The question is difficult even in limited settings such as when a trained search specialist is trying to find documents within a specialized collection (e.g., medical literature) where professionals use “a controlled style and vocabulary.”¹⁷⁵ With today’s Internet, everyone creates and shares not only web pages, but tweets, photos, videos, and more; thus, “the problems surrounding information retrieval exploded in scale and complexity.”¹⁷⁶ So with a search for Darwin, many might assume that the result should be for Charles Darwin. Yet, such a search might find results about businesses, places, people with the last or first name Darwin, the Darwin Awards, *and* Charles Darwin.¹⁷⁷ Wikipedia, video, map, news, Twitter, and, if on a mobile device, app results might be displayed.¹⁷⁸ Those who would force encounters and fear filter bubbles may think that a good thing has occurred with this wealth of options. But if the searcher was a creationist, she might find these options useless and even offensive. Such a searcher might believe that the results should indicate that Darwin and his theory are suspect, if not rubbish. And if someone searched for creationism the roles would switch once more. A creationist may be pleased if results showed arguments that disagree with and discredit evolution and related ideas. But if someone who

175. EASLEY & KLEINBERG, *supra* note 17, at 398.

176. *Id.*

177. *Cf. id.* (“For example, when someone issues the single-word query “Cornell,” a search engine doesn’t have very much to go on. Did the searcher want information about the university? The university’s hockey team? The Lab of Ornithology run by the university? Cornell College in Iowa? The Nobel-Prize-winning physicist Eric Cornell?”). A good friend has a child named Darwin, which may be rare but is still a possible search interest.

178. See Sarah Perez, *Google’s App Search Results Now Look More Like an App Store*, TECHCRUNCH (Aug. 28, 2015), <http://techcrunch.com/2015/08/28/googles-app-search-results-now-look-more-like-an-app-store/> (describing the advent of app indexing: “[I]f you were searching for a particular recipe, Google might link you to an app that offered a recipe that matched your search term. Thanks to app indexing, it could also link you directly to the page within the app where the recipe was located, after you downloaded and opened the app for the first time.”).

disagreed with creationism were the searcher, how would she make use of the results? She may simply want to see what creationists think and be happy with such results. Others may expect the results to offer information opposing creationism. And so it goes.

In addition, many other contexts matter for information retrieval. If one is in Australia, results may be improved by focusing more on the city, Darwin, than the person. Or imagine an unfortunate future where the city of Darwin is hit by a Tsunami. A search would hopefully yield news on that event, and today would likely do so; but that was not always the case. Google News was created in part because immediately after the World Trade Center attacks of September 11, 2001, searches for World Trade Center were not bringing up breaking news; instead, they yielded pages from the last index of the web.¹⁷⁹ Immediacy of an event may mean some are looking for the latest news while others would rather see more general information about that same news event. As David Easley and Jon Kleinberg put it, “[t]he same ranking of search results can’t be right for everyone.”¹⁸⁰ This point takes us back to exploration and exploitation.

Given the breadth of potential desired results and that information provision often cannot give a truly precise answer, results will frequently present the possibility of serendipity and challenge. In a dynamic, increasing information world with multiple users whose tastes and needs change continually, an information provider cannot afford to remain static. It must engage in continual exploration of options and at the same time exploit what it knows about users in general, as well as a given user both at that moment and in that specific context. By so doing, that provider can serve the specific user, and the provider can learn more about how better to offer what is the most useful information to that user, and other similar users, in the future. That said, so far it is quite difficult to be all things to all people online for news and in general.

B. Information Provider Tensions: Desires for Walled Gardens

If one site truly were a vortex from which users could not escape, that site would seem poised to create not bubbles, but walls for news and much more; and yet that is not happening. As Tim Wu noted, the walled garden was the dream of the AOL-Time Warner merger, which

179. Mark Glaser, *Google News to Publishers: Let’s Make Love Not War*, MEDIASHIFT (Feb. 4, 2010), <http://mediashift.org/2010/02/google-news-to-publishers-lets-make-love-not-war035>.

180. EASLEY & KLEINBERG, *supra* note 17, at 398.

failed in the face of the emerging public-consumption Internet.¹⁸¹ Even if one believes that Amazon's shopping and consumption recommendations, Apple's iWorld, Facebook's social network (including its News Feed, video, mobile search assistant, advertisements, etc.), Google's search and other services (including YouTube, Gmail, Maps, advertisements, etc.), Yelp's local search and app service, and so on want to keep users in a walled garden, and in some cases be all things to all people as AOL-Time Warner sought to do, that strategy has failed so far.

That these companies seek to obtain and retain users within their walls is clear and perhaps logical, but how successful they are has varied. For example, Apple's iTunes and App Store and related iPhone and iPad ecosystems are quite locked down and even gained traction as the default interface for cars' sound systems, but the advent of Amazon, Google, Rhapsody, and Spotify have loosened that hold. As options arise, consumers are able to use Bluetooth interfaces to listen to music as they want in cars, on stereos, and through other mediums. Similarly, despite Google's several attempts at services beyond search (e.g., Blogger, Orkut, Google News, YouTube, Google Docs, Jaiku (a microblog similar to Twitter), Friend Connect, Lively, SearchWiki, Wave, Buzz, and Google+), and its attempts to weave those offerings into one large, successful Google service, Google's success has been inconsistent, and it has failed in several cases.¹⁸²

These shifts track the idea that even in possibly monopolistic markets, leapfrogging—offering a “sufficiently better product” or “reduction in cost” to gain market share from an incumbent or both—is possible.¹⁸³ According to some, a firm must innovate precisely because of the threat of leapfrogging.¹⁸⁴ Indeed, several companies are challenging the very services that seemed entrenched. As of this Essay's publication, Facebook and Twitter are increasingly important as

181. WU, *supra* note 139, at 265.

182. See, e.g., Garrett Sloane, *A Look Back at Google's History of Social Media Failures*, ADWEEK (July 1, 2014, 4:02 PM), <http://www.adweek.com/news/technology/look-back-googles-history-social-media-failures-158700>; Zachary Sinderman, *Google's Long History of Social Media Attempts*, MASHABLE (July 9, 2010), <http://mashable.com/2010/07/09/google-social-media-attempts/#ZEoQPAXbwizq>.

183. Cf. Howard A. Shelanski & J. Gregory Sidak, *Antitrust Divestiture in Network Industries*, 68 U. CHI. L. REV. 1, 9, 14 (2001) (“It would ignore recent economic history to presume that Microsoft is immune from being leapfrogged and displaced from its dominant market position.”); see also Michael H. Knight & Nicholas A. Widnell, *Dark Clouds in the Distance? Network Effects and the Approaching B2B Storm*, 9 GEO. MASON L. REV. 599, 619 (2001) (noting Sega gained market share from Nintendo “in some measure from Sega's leapfrog technology,” namely, its “better processor”).

184. Cf. Shelanski & Sidak, *supra* note 183.

ways that people find news with Facebook arguably displacing Google News as the largest feed to news outlets.¹⁸⁵ Facebook has also made a big move to eat into YouTube's first-place position for online video sharing and watching.¹⁸⁶ Users can now upload videos directly to Facebook, thereby bypassing the slow downs that go with uploading from YouTube.¹⁸⁷ Thus, by April 2015, Facebook had gone from 1 billion views a day a year prior to 4 billion views a day—about half of what YouTube experiences, but a big move, nonetheless.¹⁸⁸ The move has been so successful that as of November 2015, Facebook had hit 8 billion video views a day under its method of counting three-second views.¹⁸⁹ The company also plans to develop new ways to share advertising revenue with users and track creators' intellectual property as a way to further its hold on online video sharing.¹⁹⁰ If one considers that Vine is growing and Snapchat and Yahoo! are also expanding their video options, YouTube will have to work hard not to lose its place.¹⁹¹ Apple, Citymapper, and Uber now threaten Google Maps, as they each built their own mapping services. And Google is dismantling Google+, which leaves the question of how Google will adapt and offer its many areas of service, let alone who may challenge Facebook.

There may be several possible explanations for this state of affairs, including brand issues (people may not look to a company that started in one area as a provider for other services), early mover advantage (e.g., Facebook for social, Google for search, Yelp in local search), market segments that are more separate than simply being online, and network effects.¹⁹² Although I think all have some merit, I offer that the difference between open and closed networks has much to do with these outcomes and explains why filter bubbles are not as real or possible as some argue.¹⁹³ As Wu observed about the AOL-Time Warner merger

185. See, e.g., Matthew Ingram, *Facebook Has Taken Over From Google as a Traffic Source for News*, FORTUNE (Aug. 18, 2015, 4:41 PM), http://fortune.com/2015/08/18/facebook-google/?xid=gn_editorspicks&google_editors_picks=true (“Facebook is no longer vying with Google but has overtaken it by a significant amount”). As a side note, I came across this article based on a Google News editor pick.

186. See David Holmes, *Are Facebook's Algorithms Killing YouTube?*, PANDO (July 9, 2015), <https://pando.com/2015/07/09/are-facebooks-algorithms-killing-youtube>.

187. See *id.*

188. See *id.*

189. See Josh Constantine, *Facebook Hits 8 Billion Video Views, Doubling from 4 Billion in April*, TECHCRUNCH (Nov. 4, 2015), <http://techcrunch.com/2015/11/04/facebook-video-views/#.ef3zqj2:lbYI>.

190. See *id.*

191. See *id.*

192. Shelanski & Sidak, *supra* note 183, at 6–7.

193. WU, *supra* note 139, at 265–66.

failure, the dream might have been to “direct” and “expose” users “to Time Warner’s offerings: its content, its cable TV, and its Internet services.”¹⁹⁴ That effort hoped to generate “a feedback loop, creating even more AOL subscribers,” but that was not possible.¹⁹⁵ In an open, net-neutral Internet, you can “go wherever [you] wish[] . . . [and with] search engines and domain names . . . the whole world of the Internet opens up before you. . . . At most, AOL could recommend Time Warner content to anyone logging on.”¹⁹⁶ These points relate to how information in networks can cause what are called cascade or herd behaviors and also why information in networks helps stop such events.

C. Open Networks and Popping Bubbles

The fear of information myopia, filter bubbles, or lack of access to information about competitors’ goods or services may be understood as the problems of an information cascade, but those who adhere to filter bubble logic misunderstand what generates the behavior and how strong the behavior is.¹⁹⁷ The error is in not seeing that cascades are “fragile” and can be broken if “slightly superior information” is available.¹⁹⁸ Open networks preserve the possibility of this very information being available.

A cascade occurs

when people can see what others do but not what they know, there is an initial period when people rely on their own private information, but as time goes on, the population can tip into a situation where people—still behaving fully rationally—begin ignoring their own information and following the crowd.¹⁹⁹

The problem is that cascades can be wrong, and it can take very little

194. *Id.* at 266.

195. *Id.*

196. *Id.*

197. See SUNSTEIN, *supra* note 37, at 80–84; accord Hunter, *supra* note 4, at 652–53 (“Falsehoods, though quickly propagated by the Internet, are typically remedied by contrary information spreading as speedily through the same channels.”).

198. Sushil Bikhchandani et al., *Learning from the Behavior of Others: Conformity, Fads, and Informational Exchange*, 12 J. ECON. PERSP. 151, 157–58 (1998); see also EASLEY & KLEINBERG, *supra* note 17, at 503. Other work on cascades and herds holds that a cascade requires that people ignore their private signal and later decision makers learn nothing; comparatively, in herds decision makers “become more and more likely to imitate but their actions still may provide information.” Boğaçhan Çelen & Shachar Kariv, *Distinguishing Informational Cascades from Herd Behavior in the Laboratory*, 94 AM. ECON. REV. 484, 485 (2004). In this view, herds are “fragile in the sense that a strong signal may cause behavior to shift suddenly and dramatically,” whereas in cascades “no signal can cause a change in the pattern of behavior.” *Id.* For this Essay, I follow Easley and Kleinberg regarding the point that cascades can be broken as well as herds.

199. EASLEY & KLEINBERG, *supra* note 17, at 503.

information for one to start. The claim that filter bubbles exist and will persist can be thought of as a cascade problem. In that view, someone has entered a bubble and information intermediaries simply feed the same results over and over to the user and seal out new information. Thus a way to understand the filter bubble idea is as a concern that cascades decrease learning and lower related welfare, and that they are quite difficult to stop. Although cascades can hinder learning, work on whether they last, and for how long, shows that cascades should be difficult to maintain.

The ability to connect and share information online is vital to preventing bubbles or cascades from lasting. It takes only a little difference—"the arrival of better informed individuals, the release of new public information, and shifts in the underlying value of adoption versus rejection"—to break a cascade, and as such cascades are "fragile with respects to small shocks."²⁰⁰ One might argue that a Facebook user sees nothing other than what Facebook dictates or a Google user is fed only what Google wishes the user to see. Yet, as discussed, such perfection is difficult, if not impossible to achieve. As with the AOL-Time Warner example, users on an open Internet may face suggestions—one might say strong ones—to stay with Facebook or Google services and content, but maintaining that strategy is difficult, and perhaps unwise.²⁰¹ Users change preferences. Sometimes boredom or curiosity dictates a change. Other times users find the "underlying value of [staying with a service dictates] rejection."²⁰² When that happens, some offer quick, cutting commentary about how a service has failed to meet expectations. And they can—and do—look elsewhere. They connect with others outside of these systems through other online services, mobile platforms, and word of mouth.²⁰³ In general, as people share information from outside a given network or information arena back into the system, the run in one direction is slowed down, if not

200. Bikhchandani et al., *supra* note 198, at 157–58; accord EASLEY & KLIENBERG, *supra* note 17, at 430 ("It was easy for a fresh infusion of new information to overturn [the cascade]. This is the essential fragility of information cascades: even after they have persisted for a long time, they can be overturned with comparatively little effort.").

201. WU, *supra* note 139, at 265–66.

202. Bikhchandani et al., *supra* note 198, at 157.

203. Accord Hunter, *supra* note 4, at 640 ("The Net, even the perfectly filtered Net, does not remove us from the real, physical world, where all of our usual freedoms exist without reference to cyberspace. . . . Sunstein forgets about the influence of external sources on our filtering of media. These external sources have a fundamental effect on our media choices. [People] learn about good films, books, and magazines from colleagues, billboards, free city papers, overheard discussions on the bus, and so on.").

stopped, and “the welfare of later individuals generally rises.”²⁰⁴

The more open and varied the sources of information, the more likely society can break through personal and higher-level filters to assess a question and act. Returning to news, there is a long history of trying to control news to shape society. Those efforts are less powerful and less successful with an open Internet. When the United States limited the amount of information about the second Gulf War available to U.S. news agencies, one might have expected that citizens would find that the only news available reinforced the views of the U.S. government. There was in essence one signal. The open Internet, however, allowed those dissatisfied with one signal to seek non-U.S. news sources and share them. A study of the rise and fall of three related pieces of copyright legislation shows that at least in one polarized political question, what started as single user or fringe discussions reached mainstream news outlets and allowed “citizens [to] come together to overcome some of the best-funded, best-connected lobbies in Washington, DC.”²⁰⁵ Again it takes only a few sources to break a possible cascade. But there is no guarantee that new action will occur.

CONCLUSION: IMPLICATIONS AND LIMITS

The desire for an information provider to offer relevant results makes sense; the claim that there is one answer about what is relevant does not. Information is inert, and yet grows at an amazing rate. That is perhaps why services that help us sort information continue to be created. Services that help us find the information we want are welcome, but those services cannot tell us what to do with what they offer and what we find. What we do with information is up to us. In some cases we may give an information provider specific details about our interests so that results conform to our tastes. We may also allow a provider to infer interests and suggest results based on a large range of signals about us. Along the way, information providers must engage in continual exploration of new options so that they may learn whether a better offering will lead to greater success. Whatever the way information is provided, when services fail to provide useful information, the question becomes whether we have options to help each of our respective information needs. Openness and options cannot, however, solve all

204. Lones Smith & Peter Norman Sørensen, *Observational Learning*, in THE NEW PALGRAVE DICTIONARY OF ECONOMICS 12 (Steven N. Durlauf & Lawrence E. Blume eds., 2011), <http://lonesmith.com/sites/default/files/observational-learning.pdf>.

205. Benkler et al., *supra* note 103, at 39–46 (2015).

problems related to the provision and use of information.²⁰⁶

Individuals need room to explore and exploit information on their terms. Returning to Bush's article, he dreamt of information provision evolving so that each individual could have a way to organize and use information as she sees fit.²⁰⁷ Thus, although beyond the scope of this Essay, a few points are important for information providers going forward. The claims and concerns about relevance, bubbles, and even competition lack foundation right now. But if an information provider can truly create a walled garden, or the law fosters such a garden (such as by creating a state-sanctioned monopoly or oligopoly), options and the room for personal exploration would be less—and possibly not—available.²⁰⁸ As such, the privacy issues that travel with the types and amount of data needed and used to provide ever more personal and specific results for information provision are a starting point for the debate to come.²⁰⁹

The larger debate is about what welfare means and what system society wants overall. Thus, Julie Cohen's call for "semantic discontinuity" in the networked world is an explicit argument against a seamless information experience as a way to enable what she calls the play of everyday practice.²¹⁰ I take that call to be one for greater room for exploration by users. In that sense, as Cohen has explained, individuation has a dark side, and its importance can be overstated.²¹¹ According to Cohen we need a "right to avoid individualized treatment," and we may require "interventions aimed at preserving the commercial, technical, and spatial disconnects that separate contexts from one another."²¹² And as Pasquale has admitted, he is no longer looking at these issues as ones of innovation and competition, but instead seeks to explain when an information provider is an "essential cultural and political facility."²¹³ As such, although information

206. For a detailed investigation of tensions and problems of today's networked society and why openness as commonly understood is not a cure all, see generally COHEN, *supra* note 40.

207. See generally Bush, *supra* note 1.

208. See COHEN, *supra* note 40, at 219–20 ("Networked space can be a space of dystopian domination or a space that affords breathing room for situated creativity and critical identity practice, depending significantly on the nature of its system of boundaries and permissions. Network-neutrality mandates, however they are crafted, simply do not speak to that question.").

209. See *supra* notes 97–100 and accompanying text (discussing privacy and individuation).

210. See COHEN, *supra* note 40, at 223–24 (calling for a system that allows users greater ability to work outside an information system's rules, thereby enabling engagement in creative actions not bounded by the system's rules).

211. *Id.* at 219–20.

212. *Id.* at 252.

213. Pasquale, *supra* note 6, at 172 (arguing that certain intermediaries should be regulated based not on economic terms but on "cultural, reputational, and political" terms).

providers try to tailor information provision to each user, claims that a given service is an unalloyed good may fall short on grounds that are outside the economic ones common to debates about digital media.

Insofar as regulators are reaching beyond neoliberal economics and current views of innovation, they may be doing so, because they fail on those grounds. As this Essay and others have shown, organizations compete continually and must account for changes in online behaviors, tastes, and more by exploring options and new services if an organization wishes to maintain its position, let alone grow. On that view, there does not appear to be the market failures with which this Essay began. Nonetheless, for some, the interests and rules for information provision to date are not satisfying.²¹⁴ Scholars and regulators who are questioning the core of information provision practices, may be saying something that information providers miss. If so, a key question for regulators to answer is what theory or grounding identifies harms, explains specific demands of and actions against information providers, and can guide providers going forward. This need may be understood as revisiting the Green Paper process that set policy in the early days of the consumer Internet. Alternatively, it could be an issue for legislative action and process to determine what new contours, if any, are needed.

Regardless of the path, the challenge going forward is to gather the correct facts about how information provision works and what harms are at stake from a given information provision practice, and, perhaps most importantly, to gain accurate understandings of the computer science behind these practices.²¹⁵ This last point is vital, and this Essay hopes to be part of fueling that discussion.

Thus, as this Essay has argued, the debate over filter bubbles and online competition has missed how and why the exploration-exploitation dynamic plays a powerful role in, and is currently a deep part of, how information providers operate. An information provider must explore new options (i.e., offer some difference to users) lest the provider miss the chance to learn about long-term outcomes and possible disruptions newcomers might offer. But users must also explore. If the facts show that users are complacent or choose not to engage with ideas even when they cross their paths, that says something about the nature of our democracy and society well beyond information provision concerns. If, however, a system truly limits and shuts down a

214. See COHEN, *supra* note 40, at 223.

215. Cf. David Lazer, *The Rise of the Social Algorithm*, 348 SCI. 1090, 1091 (2015) (noting the need to study the “interplay of social and computational code”).

user's ability to explore and learn, and in that sense make real use of the record, there is a real problem.²¹⁶ As far as news and shopping information go, that case does not seem to be made. Still, the criticisms and motives behind them present a challenge, or at least a reminder. If “[o]ne of the most important problems in machine learning—and life—is the exploration-exploitation dilemma,” then information providers are being told that they need to maintain a better balance between the two, not only for them but for those that they should serve: us, the information users.²¹⁷ Exactly how that will happen and what the next generation of information provision will be is the great challenge that lies ahead. It is a challenge that will require further exploration and possibly exploitation of what we have learned about information provision.

216. For example, Zeynep Tufekci has discussed a study by Facebook researchers of Facebook's News Feed algorithm and noted that the study shows that Facebook's design of the algorithm affects things such as what stories receive more clicks. Zeynep Tufekci, *How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks*, MEDIUM (May 7, 2015), <https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab#.qjwgfh e74>. She takes that outcome to show that the design could allow a user to signal she wants “more content [she'd] likely disagree with,” but right now that option is not present. *Id.* This view raises the questions of whether one can alter his or her experience with a given information provider and how a provider provides information; but the points do not explain why a given provider must offer challenging views. In other words, the critique seems to be that Facebook should allow filtering of a different kind, one with more user input, rather than no filtering at all.

217. DOMINGOS, *supra* note 2, at 129.