INFORMATION, COMMUNICATION & SOCIETY, 2017
VOL. 20, NO. 1, 63–80
http://dx.doi.org/10.1080/1369118X.2016.1199721

Routledge
Taylor & Francis Group

Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem

Tarleton Gillespie[a,b]

[a]Microsoft Research New England, Cambridge, MA, USA; [b]Department of Communication, Cornell University, Ithaca, NY, USA

ABSTRACT
Because information algorithms make judgments that can have powerful consequences, those interested in having their information selected will orient themselves toward these algorithmic systems, making themselves *algorithmically recognizable*, in the hopes that they will be amplified by them. Examining this interplay, between information intermediaries and those trying to be seen by them, connects the study of algorithmic systems to long-standing concerns about the power of intermediaries – not an algorithmic power, uniquely, but the power to grant visibility and certify meaning, and the challenge of discerning who to grant it to and why. Here, I consider Dan Savage's attempt to redefine the name of U.S. Senator Rick Santorum, a tactical intervention that topped Google's search results for nearly a decade, and then mysteriously dropped during the 2012 Republican nominations. Changes made to Google's algorithm at the time may explain the drop; here, they help to reveal the kind of implicitly political distinctions search engines must invariably make, between genuine patterns of participation and tactical efforts to approximate them.

ARTICLE HISTORY
Received 16 October 2015
Accepted 16 May 2016

KEYWORDS
Algorithm; public; search engines; visibility; frothy

Recent scholarship about algorithms and their social effects asks what it means when algorithmic information systems are inserted into social processes where (human) judgment matters (Barocas, Hood, & Ziewitz, 2013; Beer, 2009; Gillespie, 2014; Kitchin, 2014; Seaver, 2013; Ziewitz, 2015. For a growing list, see Gillespie & Seaver, 2015). Whether those judgments are about what is most relevant amid a web of information, what symptoms indicate a particular disease, or which market palpitations warrant investment or panic, the kinds of associations, categorizations, and distinctions that algorithms are designed to make threaten to join, even supplant, those made by human decision-makers. Concerns about algorithmic bias and discrimination (boyd, Levy, & Marwick, 2014; Diakopoulos, 2015; Granka, 2010; Grimmelmann, 2008; Introna & Nissenbaum, 2000; Noble, 2012; Pasquale, 2015; Tufekci, 2015) represent the applied edge of this conceptual concern. If our participation in public life (Crawford, 2015; Grosser, 2014; Pariser, 2012; van Dijck, 2013), our ability to seek educational advantages or work opportunities (Pasquale, 2015), or our

CONTACT Tarleton Gillespie tarleton@microsoft.com Microsoft Research New England, Cambridge, MA 02142, USA

© 2016 Informa UK Limited, trading as Taylor & Francis Group

position in the market (Citron & Pasquale, 2014; Poon, 2007) is being determined, or at the very least adjudicated, by algorithmic systems, then we must know more about the assumptions upon which they are based, the information about us upon which they act, the priorities they serve, and the ways in which they shape, distort, or tip the process (Ananny, 2015; Bucher, 2012; Cheney-Lippold, 2011; Graham, 2005; Hallinan & Striphas, 2014; Mager, 2012; Rieder, 2012).

However, this tendency to treat algorithmic systems as new, hidden, and powerful mechanisms with built-in values can pull our work toward some predictable potholes: treating the technology as a singular object, distinct from human operators and institutional arrangements; treating the algorithm as a secret, tightly guarded and therefore protected from critique; and treating the world in which the algorithmic system operates as otherwise simple, untouched, and vulnerable to manipulation.

Grimmelmann (2014) recognizes this problem best in his analysis of the legal implications of search results. He notes that opinions about search engine liability fall into two camps, one treating the search engine as a neutral conduit delivering results to users, the other treating it as having a kind of editorial power to pick and choose. He suggests that both views are incomplete, because both overlook the user as the third point in this triangular relationship. Search engines act more like advisors to the user, responsive to the query and suggesting which results are most relevant. It is an important reminder, one that undercuts the objective/subjective binary that seems to constrain discussions of algorithmic systems and their impact.

I want to make a parallel argument to Grimmelman's, that too often we treat the information providers as independent of search engines, as if they are merely standing on the edge of the field waiting to be picked for the team. I believe that this profoundly overlooks the strategic efforts of the content providers. Precisely because information algorithms make judgments that can have powerful consequences, those interested in having their information selected as relevant will tend to orient themselves toward these algorithmic systems, to make themselves *algorithmically recognizable* (Gillespie, 2014), in the hopes of being amplified by them.

As soon as nascent search engines began to mediate users' interaction with the web, so also emerged a range of tactics for trying to be recognized by them. In his discussion of the politics of search, Grimmelmann (2008) noted some of these: Googlebombing, link farms, government intervention. But I think it goes further. Web design handbooks and etiquette guides offered tips for designing web pages to be congruent with how search engines index and judge sites as relevant. Bloggers and video creators asked readers to link back, like, retweet, or otherwise circulate that content, in the hopes of generating the signals that search engines value; search engine optimization (SEO) firms and spammers developed techniques, premised on guesses about how search algorithms worked, to boost their clients ranking and gain additional traffic. To some degree, every contribution to the public web in some way desires to be seen, which generally requires being recognized and amplified by Google.

These strategies of visibility are analogous to efforts to be recognized and amplified by other kinds of intermediaries: sending press releases to news organizations, staging events with the visual impact that television craves, or making spokespeople and soundbites available in ways convenient to journalists. 'Generators of information for the press anticipate the criteria of the gatekeepers in their efforts to get through the gate … How do you

get a piece of information to 'pass' as news?' (Schudson, 1989, p. 265). Such tactics belie the myth of a world waiting to be reported upon and journalists independently seeking the most important parts, and trouble the distinction between a genuine desire to be heard and a strategic effort to game the system. These also have much in common with tactics of avoidance and obfuscation in the face of surveillance systems (Brunton & Nissenbaum, 2015; Marx, 2009): in this case, it is the aim not to be seen, but in the same way actors will engage in 'neutralization techniques' (Marx, 2009) that must understand the workings of the surveillance system and fit themselves to the contours of its attention and inattention.

Search algorithms have a set of organizing criteria for the kind of phenomena they seek: particular kinds of websites, particular patterns of incoming links, and particular behaviors of users, all read as signals of a genuinely emergent and non-strategic demonstration of a site's true relevance. Being algorithmically recognizable, then, means simulating this particular (ideal/idealized) type of activity. And it means doing so amidst others also trying to simulate that same activity, for a variety of reasons of their own. As we all jostle for recognition, it falls to search engines not just to determine what is relevant, but to distinguish between genuine signals and 'gaming the system' – a distinction that is arbitrary but conventional.

Rather than thinking in terms of algorithms and their values, we might take a lesson from Brunton's (2013) discussion of spam and spam filters. Brunton notes that all communication is tactical; rather than a coherent genre, spam is whatever is defined away as illegitimate by the mechanisms designed to do so. Not only is Google a spam filter too, but the judgments of relevance Google must make must cope with the fact that all communication is tactical, proclaims its own value, wants to be recognized, turns to face the microphone, performs that value in social terms, and tries to appear legitimate in its eyes. Google must discern the relevant from the irrelevant, amid a shifting sea of bids to appear so. Search algorithms may have political ramifications, but our understanding of them will be richer if we see them not as having built-in values but in terms of the tactical and ad hoc ways in which they make determinations of relevance, even as what they are judging is shifting so as to benefit from that determination.

In this essay, I will look at the decade-long effort by sex columnist Dan Savage to criticize U.S. Senator Rick Santorum by popularizing an alternative definition of his name. Savage's site quickly became the top Google search result for 'Santorum' and retained that position above official biographies, news coverage, and even the Wikipedia entry on Sen. Santorum, until those results abruptly changed in 2012 during Santorum's effort to win the Republican nomination for the U.S. presidential election. This case, on one level, could be understood as a powerful example of the political values built into search algorithms. But by focusing on the tactical moves made by Savage, Santorum, and Google, I hope to instead highlight (1) the way public actors attempt to make themselves recognizable to information intermediaries, and (2) the way information intermediaries, algorithmic or otherwise, are forced to discern between acceptable and unacceptable efforts to be seen, when being seen has political currency.

Spreading it around

In 2003, then U.S. Senator Rick Santorum shocked even the AP reporter interviewing him when, in a comment defending laws that curtail private sexual acts,[1] he equated

homosexuality with adultery, polygamy, and incest, and then with 'man on child, man on dog, or whatever,' all behaviors he felt 'undermine the basic tenets of our society and the family.' Like many others, sex columnist Dan Savage took umbrage with Santorum's remarks; unlike many others, he wrote an angry op-ed in the *New York Times* in response.[2] That could have been the end of it, except that one of his readers proposed that Savage and his audience should respond by naming a sex act after the Senator.[3] Savage turned this idea back to his readers, and after receiving over 3000 suggestions, Savage invited readers to vote on the best,[4] then announced[5] that 'santorum' should now and forever be defined as – brace yourself – 'the frothy mixture of fecal matter and lube that is sometimes the byproduct of anal sex' (Figure 1).

At first, the act of selecting the winning definition seemed to be the entirety of Savage's political gesture. But soon after, he purchased the web domains santorum.com and spreadingsantorum.com, and posted his new definition at both. On the splash page, both the definition itself and the words 'santorum, senator, rick santorum' were included in the <meta> tags, an (old) technique for drawing the attention of search engine indexing bots. As readers, bloggers, and the press began linking to and commenting on the site, the site quickly became the top search result returned in Google to the query 'Santorum.' At least within Google's index, and for unsuspecting Google users, the Senator's name had been successfully redefined. Many surmised, though it is hard to know, that Savage's site and its prominence in Google's results were influential in Santorum's re-election loss in 2006.

Others had already proven that Google's algorithm could be manipulated. Early tricks to boost a page's ranking included loading it with popular but irrelevant meta tags, filling the bottom of the page with invisible text, and including links out to popular sites. Other tactics, that were perhaps less about 'gaming' the algorithm and more about being recognized by it, included 'web rings' and 'blog rings' where sites with shared interests would agree to link to each other, and search registration services that would alert search engines to a new site and invite them to index it. Eventually, an entire industry offering 'SEO' emerged, offering consultation on how to design websites to boost their ranking on major search engines, based on some divination of Google's evaluative criteria. And

Figure 1. Splash page, Spreadingsantorum.com. Used with permission from Dan Savage.

then there was a practice known as 'Google bombing' (Grimmelmann, 2008): get enough people to link the phrase 'miserable failure' to President George Bush's biography, and watch as the search for 'miserable failure' delivers up Bush's site as the top result.[6] (The trick, though, requires telling people to search for 'miserable failure'; it does not affect what results return to a query about Bush himself. This makes it fundamentally an inside joke.) 'Squatters' registered domain names that represented well-known names or brands, hoping to draw users searching for their favorite store or celebrity. The white supremacist group Stormfront extended this technique further, quite effectively and reprehensibly. In 1999, the group secured the domain martinlutherking.org, posting there what appeared on first glance to be an encyclopedia entry on the civil rights leader; examined more closely, it was in fact a skewed and slanderous profile of the man that betrayed the site's extremist beliefs. For years, their site was the top result in Google to the query 'Martin Luther King Jr' – the site remains in the first page of results.

What Savage and his readers did, as one critic put it, was 'calculated character assassination'[7] – calculated both as in deliberate, and as in mathematical. But it was not Googlebombing. He and his readers deliberately and successfully unseated other sites, including Santorum's official website, securing theirs as the 'legitimate' top result to the query 'santorum.' It is one thing to knowingly look up a specifically worded criticism, to see it turn up a public figure's name; it is another to have all searches for the public figure's name point, first, to a critical parody that figure. (In this sense, it is closer tactically, if not politically, to the Stormfront site.)

The campaign to algorithmically redefine Santorum's name required, first, a not insignificant number of users to link to Savage's site, and in a way that Google's algorithm would recognize as akin to linking to the 'correct' or 'most relevant' site: for instance, by making the word 'Santorum' the text anchor for their link back to spreadingsantorum.com. Google's indexing tools noted these links, and included them in their calculations as indicators of value, boosting the PageRank of Savage's site.

This also required a great deal of coordination. Google designed its search algorithms to recognize patterns that represent the separate and aggregate the links and clicks of millions of users. To trump the algorithm, Savage's coordinated effort would have to be on the scale of other kinds of less coordinated behavior, like supporters of Santorum who might often link to his campaign site. But Savage is no ordinary user. He is a public figure with a great deal of credibility, from his syndicated newspaper column and podcast, his public appearances, and his activism in the gay rights movement. He has a devoted readership, many of whom were already in on the campaign from the start. And his readership overlapped with a broader community of gay activists who already had experience in the tactics of political visibility, including online.

There is a distinction often made, between coordinated efforts to 'game' a search engine (like Googlebombing and SEO tactics) and the 'genuine' output of independent web producers and users, demonstrating the value of a site from their linking and clicking behaviors. The distinction is a false one. Most contributions to the web are somewhere in the middle, where people in some way coordinate their efforts in order to help make their content visible to a search engine, out of a 'genuine' desire for it to be seen. Activist organizations publicize their efforts and gather supporters online; companies urge their customers to post about their product, sometimes in exchange for rewards; writers send their posts to their friends or tweet them to their followers. It is not easy to distinguish, ethically or even

practically, between link spam, paying for links, encouraging readers to link, enjoying links generated by people already invested in your project, sharing the content through public networks in the hopes that it will circulate further, or achieving links organically when people stumble onto your site. In more cases than we might care to admit, search rank is a product of a combination of these; the last one, so often held up as the ideal, may be increasingly rare. Not only is there a gray area between 'genuine' linking behaviors and coordinated efforts to game the search engine, it is not even clear that the two are all that different.

Making it stick

The spreadingsantourm.com site remained the top result on Google to the query 'santorum' through the remainder of Santorum's Senate term, his failed re-election bid in 2005–2006, and on through to 2011, when Santorum announced his run for the Republican presidential nomination. After having allowed the spreadingsantorum.com site to stagnate, Savage added a blog that tracked criticism of Santorum and any coverage of the 'frothy' definition, and share buttons allowing users to 'like' the site on Facebook and Google+.

Some wondered whether Santorum's reappearance in national news coverage would help more official, legitimate content supplant Savage's site in the Google index; timely headlines in major news outlets might very well supersede a nearly decade-old site with very little new content. But the press and political humorists also helped to revive the association between Santorum and Savage's definition, and in doing so, likely strengthened it in Google's ranking. News coverage of the lengthy nomination process needed to fill countless broadcast hours and regularly refresh their sites. Reporters desperate for any topic related to the race often found the 'santorum' neologism, and then wrote about whether Savage's campaign might affect Santorum's chances. Even news coverage decrying Savage's site could become fodder for Google's calculations: even if an article did not link to Savage's site, it may spur readers to conduct a search of their own and click on Savage's site as the result. Every mention by John Stewart, Stephen Colbert, or Bill Maher on their late night programs, and there were many, provoked queries for 'santorum.' And Savage noted these on his blog, which both added new content to the site (Google factors a site's 'freshness' into its calculations[8]) and refracted those mentions back to his fans, potentially generating even more queries and links. Since the crux of Savage's critique is not the site itself, but the fact that it tops Google's results, part of the political theater for users who support Savage, or are just curious, involves enacting that theater themselves: searching for the term, seeing it appear, and (perhaps) clicking through to the site. All of this provides Google further data confirming the site's relevance.

It is easy, but misleading, to think of the web as a pile of undifferentiated sites, all independently generated and vying for attention, to which the algorithm discerns and selects the most relevant. Even the interface of Google's search engine suggests this, the way search results are lined up in marching order, ranked only by Google's judgment of their relevance to the query. This notion radically masks the fact that, while sites may be independently authored, they are coordinated by professional routines, world events, the rhythms of other media and information sources, and the quirks of circumstance. Google's search engine cannot return a result as first simply because it has the greatest

number of hits that day or the greatest number of incoming links; those may be artifacts of the rhythms of information production and public attention that surge through the web like tides pushed by the moon. The moon, in this metaphor, could be many things: election cycles that generate news stories at particular moments and with particular velocities; media events that provoke bursts of writing about the same thing at the same time; the daily and weekly cycles of information professionals like journalists; the unexpected velocity of a cultural phenomenon or viral bit of content; the deliberate efforts of industries dedicated to shaping the public information landscape, from advertisers to promoters to spin doctors to activists to search engine optimizers to spammers. The success of Savage's campaign depended not only on Savage, his readers, amused bloggers, and gay rights activists. It also benefitted from the dynamics of media cycles and election cycles, which together helped to generate recurring attention and links to Savage's site. This is not exactly coordinated activity designed to game the search engines, the way the initial tactics of Savage and his readers arguably were. But the user activity Google pays attention to is coordinated, by powerful forces.

This posed an additional distinction for Google to make: should incoming links, some from high-status news and commentary sites, further legitimate Savage's page and boost its ranking? Does a surge in attention provoked by the media represent the kind of interest search engines should reward, or counterbalance? Google designers probably have a unique understanding of the complicated and overlapping rhythms of information production online. In every case, Google must decide what to make of these rhythms, how to value them in relation to each other, and how to incorporate them into the calculations that produce search results.

Keeping it clean

As Santorum's fortunes in the 2012 Republican nomination improved, the political press paid more and more attention to him, and paid more and more attention to Savage's site and its prominence in Google's search results. Or to say it another way, Santorum's public visibility, driven by his own strategic campaign efforts and the efforts of his political supporters, and amplified by the media cycles attendant to the U.S. presidential race, challenged the visibility of Savage's campaign as well.

Reporters began to ask Santorum if he would request that Google remove the site from its index. Santorum gave three different answers over the course of 2011, demonstrating a range of available positions one might regarding the nature of Google's algorithm. First, he said he would not ask Google to remove the site, a position that either sees Google's algorithm as beyond intervention, or sees intervention as an ineffective or politically risky move. Drawing attention to Savage's site might even strengthen its ranking by leading even more people to it. Later, he lambasted both Savage and Google in a fundraising letter – leaving Google's algorithm intact, but finding a different kind of (financial) value in it. Finally, in September 2011, he asked Google to remove Savage's site from their index. As he put it, 'If you're a responsible business, you don't let things like that happen in your business that have an impact on the country.'[9] This suggestion, that Google shares some responsibility for how their search results might have political impact, was paired with a dig at Google's possible bias as a company – 'I suspect if something was up

there like that about Joe Biden, they'd get rid of it'[10] – akin to the charge of 'liberal bias' so often leveled at broadcast media.

Google was then faced with a choice: remove Savage's site, or let it stand in the rankings. Either decision would be a political one, and would be perceived as such. Over its history, Google has been nearly unwavering in its stance that its search results should not be altered or censored. Famously, in 2004, Google refused to alter the index when the hateful, anti-Semitic site JewWatch ranked at the top of the search results for the term 'Jew' (Grimmelmann, 2008).But this stance is not without exceptions. Google removes spam sites, sites charged with being defamatory, and sites challenged as copyright infringement under the Digital Millennium Copyright Act (DMCA).[11] On several occasions, they have temporarily demoted commercial sites, including J.C. Penney[12] and Overstock[13] for optimizing their sites in ways Google deemed unacceptable. And they have, on occasion, removed content for being offensive. In 2009, a racist image of Michelle Obama turned up as the top result for her name on Google's image search. In response to criticism, Google first refused to remove it. But after continued criticism, Google delisted the image from the index, indicating on the results page that it had done so. (They later were able to remove the image from the source, as it happened to be on a blog hosted on Blogger, a Google-owned site.) Recently, Google began to voluntarily remove links to revenge porn. And the new European 'right to be forgotten' rule, Google and all search engines must remove specifically requested links pointing to 'inaccurate, inadequate or no longer relevant' pages from the search results for a given person's name.

In a public response to Santorum's request, Google refused to alter the index. 'Google's search results are a reflection of the content and information that is available on the Web. Users who want content removed from the Internet should contact the webmaster of the page directly.'[14] A Google spokesperson did note that Google does not 'remove content from our search results, except in very limited cases such as illegal content and violations of our webmaster guidelines.' Or, 'Search engines pride themselves on being automated, except when they aren't' (Grimmelmann, 2008, p. 950). Google's statement carefully figures the algorithm as unmanaged, while also leaving room for the fact that Google does alter the index, not just under legal obligation, but under specific guidelines – guidelines that they crafted, which means of course that they could alter the index in this case if they so chose. They also noted that Savage's 'spreadingsantorum' site *was* already blocked, at least for some users: Google's SafeSearch function already prevented Savage's site from turning up for users in 'restricted' mode. Google noted that it uses algorithmic methods to identify sites that should be restricted by SafeSearch.[15] This will be important in a moment.

Perhaps the most intriguing aspect of Google's response was a separate comment made by their head of global communications, who noted that, 'There definitely are people who are finding this to be the best answer to their question, and they are indicating this by either clicking on this result or linking to this result as the best answer to that question.'[16] We could take this comment two ways. Either he means that Google does not really care about meaning, it cares about user satisfaction: if more users querying 'santorum' click Savage's link, then in Google's estimation, it is the more relevant link, because it is what users chose. The site's rank in the results is confirmed by users selecting it, which justifies its rank. Or, more radically, he is reminding us that some users entering the query 'santorum' *are* in fact looking for Savage's definition; for those users, the meaning

of 'santorum' *is* 'a frothy mixture … ' and Savage's site *is* the correct response to the query. For some users, the term has already been redefined and, coming full circle, Google's results confirm it. In this loop, Google is responding to the popular sense of meaning, indicative of it, constitutive of it, and proof of it, all at once.

Something different in the back end

However, not long after Google asserted that their algorithm is impartial, sacrosanct, and maybe even correct, something did change. In late February 2012, search engine watchers noticed that Savage's site had dropped in Google's rankings for the search 'santorum,' so much so that it no longer made the first page of results. This precipitous drop did not happen on competitors Bing or Yahoo (at least at first). It was a small victory for Santorum, perhaps, though a largely pyrrhic one: curiously, the Urban Dictionary definition that explains (and repeats verbatim) Savage's neologism had claimed the first place on Google, above Santorum's own campaign site and Wikipedia's entry; and the blog run by Savage within spreadingdsantorum.com appeared in the fourth position.

Google has not been particularly forthcoming about the change. Some surmised that Google had caved to political pressure – a reminder that, no matter why the change occurred, it is extremely difficult for intermediaries to avoid the charge of censorship (Gillespie, 2012). Few expert commentators, however, believed that Google had directly manipulated Savage's ranking. As Danny Sullivan of Searchengineland.com noted,

> To date, Google has refused to make any change specifically to the listing, which is pretty much in keeping with how it approaches these types of issues. Instead, Google prefers to resolve tricky issues like these by looking for solutions that may impact a wide range (of) listings.

Sullivan found it telling the splash page was not gone from the index altogether, which might suggest a complete removal. He also noted that, just the day before, Google had announced a bundle of changes to its algorithm.

Google has made hundreds of updates to its search algorithm over the years, a fact that should (but for some reason does not) undermine the veneer of impartiality it continues to enjoy. SEOmoz counted 92 major updates between 2000 and October 5, 2012,[17] and claims that Google makes 500–600 smaller changes every year. These adjustments represent Google engineers' effort to deliver ever more 'relevant' results, and to stay ahead of spammers, content farms, and SEO tactics. Such updates are in fact many changes bundled together, each of which has been tested in-house and, in some cases, in 'A/B' tests[18] where thousands of users unwittingly use a version of Google's search algorithm with the proposed changes, and their usage is compared to those still using the previous version. (Between ongoing A/B testing and the personalization and localization of results, not only do 'we' not get the same results from Google's search engine, but it is hard to say that there even *is* a single search engine. Google is making available many, overlapping algorithms that only appear to be a single tool.)

When Google changes the criteria by which their algorithm determines relevance in one of these major upgrades, almost by definition, some sites rise in the ranking and others drop, sometimes precipitously. These are hard to pinpoint, however, because changes to the algorithm are not always announced publicly, and because shifts in results can

occur for other, exogenous reasons. Nevertheless, businesses can go bankrupt when Google changes their algorithm (Battelle, 2005) Google is keenly aware of this, though they are adamant that these effects are not a reason not to make these changes – they are precisely proof that changes should be made. Anticipating criticism from affected sites, Google commented on a significant update in 2011:

> We can't make a major improvement without affecting rankings for many sites. It has to be that some sites will go up and some will go down. Google depends on the high-quality content created by wonderful websites around the world, and we do have a responsibility to encourage a healthy web ecosystem. Therefore, it is important for high-quality sites to be rewarded, and that's exactly what this change does.[19]

Changes to the algorithm are always positioned as new progress toward an unchanging goal.

Could the update explain the sudden and precipitous drop in Savage's site? The timing certainly suggests it, but pinpointing the change responsible is less simple. The February 2012 update included 40 distinct changes to the algorithm, the most Google had ever rolled into a single upgrade.[20] Sullivan identified two changes that might have inadvertently demoted Savage's site. First, Google claimed to have improved their techniques for identifying 'official' pages and ranking them more highly than others. Second, Google had adjusted the Safesearch algorithm, hoping to improve the identification of 'adult' content and more effectively exclude it from searches that are not adult in nature. (In other words, when you search for the term 'toys,' the algorithm should not return links to sex toys, even if they otherwise rank highly for the term.) After some hypothesizing, Sullivan was contacted by a Google representative, who confirmed that it was the changes to SafeSearch that had demoted Savage's 'spreadingsantorum' site.[21]

Still, this is not an entirely satisfying explanation. Urban Dictionary's entry for 'santorum,' which includes Savage's definition, still ranked at or near the top after the change. With the identical words, this page is presumably just as 'adult' as Savage's, and should also have been tagged by Google's algorithms – even more so, as Urban Dictionary is an entire site dedicated to unsavory definitions, many much more 'adult' than Savage's. Urban Dictionary's entry soon also dropped in the rankings for a search on 'santorum' as well, suggesting that it had eventually also been flagged as 'adult'. But it seems odd that it would not have already been identified as such.

Others disagreed with Sullivan's theory, even after it was confirmed by Google. Rishi Lakhani, interviewed at Searchenginewatch.com, suggested that

> in general, it looks like a result of poor SEO … Google muddied the water by blaming safe search, but that appears totally untrue. They don't want people to have a potentially strong example of their new 'official page detection' (OPD) algorithm shift.[22]

According to his theory, Savage's splash page may have long been incorrectly identified as an 'official' page for Rick Santorum, an error corrected by the new upgrade. This would explain why Savage's site had ranked so highly for so long even with so little new content. But it does not fit neatly with the fact that Savage's site ranked so highly at the start, before a page's 'official' standing was even a part of Google's calculations.

A third possibility is that Savage's site was snared in Google's ongoing efforts to fight spam. In February 2011, Google introduced a major update called Panda, designed to

identify 'shallow and low-quality' sites that were enjoying higher ranking in their results than sites with more informative content. The impact of Panda, according to Google's own statement, was significant, affecting nearly 12% of results. This was way before Savage's site dropped – but Panda was not a one-time intervention: on a monthly basis, Google updated Panda, each time blocking more sites. So, it is also possible that, with the latest update to Panda, Savage's front page and the incoming links that pointed to it were interpreted as spam, and de-listed as such. This does not fit neatly, however, with the fact that the Savage splash page only dropped, but did not disappear entirely.

Other commentators had still different ideas. It is, of course, possible that Google used the update as political cover, a chance to specifically demote Savage's site without having to admit to doing so. It is even possible that Google engineers do not even know why it changed. Because changes to the algorithm are opaque to both users and critics, debating what caused Savage's site to drop requires a kind of mental reverse engineering, and cannot offer a clear, convincing, or even stable answer.

It is unlikely that Google directly manipulated this particular search result. But the implications are actually far more significant if they did not. Let us assume that Google did not demote Savage's site specifically, and did not purposefully conjure up the update just to cause it to drop. Let us assume that Google made a policy decision in 2011 to leave the index alone, despite Santorum's request; that they made the February 2012 upgrades to the algorithm in order to serve up more relevant results, in their best assessment of relevance; and that the drop in Savage's rankings is the result of either the changes to how Safesearch handles adult content for non-adult queries, or the recognition and elevation of sites deemed official, or its classification as spam. Google may not engage in manipulation, but it must make categorical and *a priori* distinctions about what kinds of results to prioritize, when, and for whom. And it must do so with an eye toward how information providers will then try to emulate these distinctions.

While the Google team responsible for improving the algorithm may have been unaware of the effect it would have on Savage's site, the changes it made were nevertheless animated by specific political presumptions: about the proper contours of quality public discourse, about the difference between rich content and shallow, about where adult content should appear, about the importance of official sources of information, about the difference between activism and spam. For each, Google must algorithmically distinguish between highly valued and 'genuine' signals of relevance, and the specious simulation of those very same signals.

Adult/non-adult: From one perspective, the quiet exclusion of adult content from searches deemed to be 'not adult' makes a certain sense. It is clear why Google would not want users to be regularly confronted by unexpected pornographic results to their innocent queries – imagine the child searching for 'barbie doll' or the cancer patient searching for 'breast exam,' startled to encounter X-rated alternatives amid their results. There are, of course, challenges in making such distinctions: what makes a site adult, and what makes a search not adult. On the other hand, search engines can count on the fact that it is far worse (publicly, ethically, and economically) to deliver 'sex toys' links to the user who asked simply for 'toys,' than it is to make the user searching for sex toys have to clarify their initial query. However, deciding that adult content should not be returned to a non-adult query also disables precisely the kind of political speech Savage was engaged in, a kind that has played an important part of Western political

discourse (Naron, 1991). One might argue that this tactic is more damaging to the public discourse than whatever valid political point it might make; many critics have said as much. But this was (likely) not Google attempting to clean up political discourse; it was a decision based on protecting users from accidental offense that, as an unintended consequence, may have also demoted one form of political rhetoric.

Official/unofficial: That Google attempts to identify 'official' pages and elevate them in the ranking also makes sense. While Google's algorithm began by calculating the relevance of a site based on incoming links, rejecting the editorial approach of Yahoo for a kind of 'wisdom of the crowd' quantified populism, a lot has changed since then. The web has grown from a collection of home pages and special interest sites to a massive information archive that millions of users count on for reliable information, trustworthy commercial transactions, and accurate news. But to algorithmically identify 'official' pages and weigh them more heavily than others is a specific intervention, one infused with a particular theory of democracy in it. To privilege official sites over unofficial ones is to amplify those official voices in the public square. To put it another way, the algorithm could be designed to do the exact opposite: it could grant 'unofficial' pages (like Savage's) higher standing, precisely because they do not have the benefits of amplification that official information sources usually do (other outlets of speech, financial resources, built-in credibility). This is not to argue that this would be the 'right' design, only that every design has a theory about quality public discourse embedded within it.

Coordinated political action/spam: Search engines must seek out and block ever more sophisticated forms of spam. This now includes what Brunton (2013) calls 'spam created by machines for machines' (p. 115), including planting links that point back to a target site in the comment spaces on blogs, entire 'link farm' websites with links back to thousands of target sites, and simulating user traffic in ways that the search engine will treat as true signals. Search engines are now in the business of discerning 'genuine' online activity (commenting, linking, and user traffic) from the spam simulation of it; and spammers are in the business of emulating the way people use the web, so as to be mistaken for genuine by search engines. Spam works by veering as close as possible to other kinds of legitimate public activity: people using the medium of communication available, to draw attention to a particular resource, by gathering up support from others to demonstrate its importance. Sounds like spam, but also savvy political organizing. As Brunton (2013) notes, 'there is always friction not around the most egregious case … but at the blurry places where spam threatens to blend into acceptable use, and fighting one might have a deleterious effect on the other' (p. 163). When Google must discern link spam from other forms of strategic and coordinated meaning making, they have an important and precarious job on their hands.

We all do it

Rather than seeing search engines as interveners in a waiting field of content, we must recognize that information producers vie for attention, and are aware that search engines can give it. Authors aim their contributions toward the available mechanisms that could amplify and certify them. Remembering that information is not raw and information producers are not passive sets up a more sophisticated story about the workings and implications of the (algorithmic) mechanisms that may act upon them. And while Google's

algorithm may be premised on the links and clicks of individuals, as if those links are unaware of Google's attention to them as signals, public expression is in fact collective and collaborative. Political actors like Savage and his readers will turn together to face these information intermediaries as one and make themselves recognizable as relevant.

These efforts to face the algorithm, which require anticipating its workings and designing contributions so as to be recognized by it, can help shed light on the algorithmic systems themselves. What they do in anticipation of algorithms tells us a great deal about what algorithms do in return. While not all stakeholders have a sophisticated and accurate understanding of how an algorithmic system actually works, often depending on ad hoc, inaccurate, and outdated lay theories (Sandvig, Hamilton, Karahalios, & Langbort, 2014), their tactics can nonetheless be revealing of how an algorithmic system works, how it is imagined to work, and how users believe it *should* work.

So, we cannot simply study algorithms and their effects; rather, we must study the interaction between providers of information and algorithmic assessors of information, sometimes a confluence of interests and sometimes a contest, and the results that these interacting forces generate. And the jostling of interested voices that search engines encounter are themselves bound up in complex cycles of production motivated by other intermediaries: the cycles of traditional media, the rhythms of political campaigns, the disruptions of external events. These cycles are themselves sometimes fueled by highly motivated and organized efforts to gain public visibility by particular stakeholders.

This offers a strong reminder that, while algorithms may introduce some unique dynamics into how information is evaluated, specific to their particular computational, automated, and data-centric affordances, these are in another sense just media. They are socio-technical institutions that generate, select, and circulate public information – they mediate – and so are analogous in many ways to all the media that predate algorithms and even computation: publishers, libraries, broadcasters. By and large, questions of the power of algorithms to discern and amplify some information over others are (and should be) the same questions we have asked for decades, about the power of intermediaries.

One might argue, in fact, that Santorum's 'man-on-dog' statement about homosexuality, the one that angered Savage and his listeners and started this whole business, was much the same tactic. It was a deliberately constructed effort to express a political opinion in such a way that it would be amplified by the medium to which it was addressed – in this case, television and newspapers – so that, with the prominence and legitimacy afforded to it, it would thereby shape the debate. So while Savage may have been trying to game Google's algorithm so that an outrageous statement could enjoy more visibility than it deserved, Santorum's sound bite might have been similarly gaming the AP and cable news: a 'broadcast meme', or 'broadcast spam,' similarly designed to take advantage of a system that circulates such things.

When we are concerned about the power of information intermediaries, algorithmic or otherwise, we are fundamentally concerned about questions of visibility and meaning. This is nowhere more apparent than for search engines, which grant order web pages not just from first to last but, in doing so, from most visible to least. And in doing so, search engines stabilize meanings by offering up certain results in response to particular queries. As Thompson (2005) notes, 'Mediated visibility is not just a vehicle through which aspects of social and political life are brought to the attention of others: it has become a principal means by which social and political struggles are articulated and

carried out' (p. 49). This was a question for traditional media long before it was one for search engines. But again, visibility is not simply granted by the search engine: visibility is a prize that is actively sought and sometimes vigorously fought over. The results of that tussle, both between competing information providers and between those information providers and the search engine, results in the results.

Politics too is often a struggle for visibility. Any community trying to enact political change must struggle in some sense to be visible on the public stage. For those wishing for their political perspectives to be heard, those mechanisms that grant or withhold public visibility are of signal importance (Gitlin, 1980). Visibility generates audiences, and the anointing of a particular perspective as relevant grants it public legitimacy. Politics is also often a struggle over meaning. While Savage's site may be an extreme case, where what is being contested is the meaning of a man's own name, it is easy to point to more reasonable examples: much of the political debate over climate change, or universal health care, or gay marriage, or corporate speech, involves contested definitions of the terms themselves. To some degree, political campaigning is in part a semiotic undertaking, designed to confer and stabilize a set of meanings attached to a candidate's name: on his 2016 campaign website, Santorum's 'about' page begins 'Rick Santorum is a conservative committed to restoring the American Dream for hardworking Americans.'[23] This is just as much of a semiotic assertion as 'Santorum: the frothy mixture … ' is. But we see this one as more legitimate because Santorum is defining himself – or, in fact, a network of campaign advisors and staffers with his guidance, supported by thousands of volunteers who help his campaign in part by restating that definition in their own public contributions. Santorum's Republican rivals also tried to 'define' Santorum, as too conservative or unprepared for office or not as likely to win a general election; these too were efforts to associate a meaning with a name, and they too depended on information intermediaries to amplify and certify those meanings. It is left to Google and other information intermediaries to decide whether each of these semantic, collaborative, and tactical affirmations are genuine, or just designed to assert one interpretation over the rest, and whether to pick up on and certify those signals as legitimate or to downplay those meanings as mere political gamesmanship.

Savage wanted public visibility for his message, a message in which he challenged the meaning of a term. He had the resources and the expertise to mobilize people to help him make this contested political claim, and found a clever strategy for amplifying that collective political expression by making it 'algorithmically recognizable' to Google's search engine. This required flirting with the obscene, giving an unofficial site the trappings of an official one (or a carnivalesque version of it), and getting dangerously close to being spam. Google and other search engines are in a position of power, not because they are algorithmic, but because they sit in a uniquely powerful position where they get to grant visibility and certify meanings.

Those who distribute public discourse must draw distinctions between that which we want to avoid, even when it tries to look like what we want to protect, and what we want to protect, even when it looks a lot like what we want to avoid. The greater the demand for making that distinction, either from users or critics or whoever pays the bills, the greater the risk that will befall those who end up on the wrong side of that distinction. How the distinction is made, both technically and institutionally is an important cultural and political question. And it is a question we have grappled with around

information intermediaries long before the introduction of computers and algorithms. These are well-worn political dilemmas, about the role of intermediaries doing their business among competing cultural and political interests, leaning on and reifying categories of distinction, with consequences for what is visible and what is not, and for stabilizing contested political meanings.

Notes

1. Associated Press. (2003, April 2003). Except from Santorum interview. Retrieved from http://usatoday30.usatoday.com/news/washington/2003-04-23-santorum-excerpt_x.htm.
2. Dan Savage. (2003, April 25). G.O.P. Hypocrisy. *New York Times*. Retrieved from http://www.nytimes.com/2003/04/25/opinion/25SAVA.html.
3. Dan Savage. (2003, May 15). *Savage Love*: Bill, Ashton, Rick. *The Strangler*. Retrieved from http:://www.thestranger.com/seattle/SavageLove?oid=14267.
4. Dan Savage. (2003, May 29). *Savage Love*: Do the Santorum. *The Strangler*. Retrieved from http://www.thestranger.com/seattle/SavageLove?oid=14422.
5. Dan Savage. (2003, June 12). *Savage Love*: Gas Huffer. *The Strangler*. Retrieved from http://www.thestranger.com/seattle/SavageLove?oid=14566.
6. Marissa Mayer. (2005, September 16). Googlebombing "failure". *Google Official Blog*. Retrieved from https://googleblog.blogspot.com/2005/09/googlebombing-failure.html.
7. "Pro-Family" activist calls on Dan Savage to fix Rick Santorum's "Google problem". (2012, January 5). *Huffington Post*. Retrieved from http://www.huffingtonpost.com/2012/01/05/dan-savage-rick-santorum-google_n_1186784.html.
8. Amit Singhal. (2011, November 3). Giving you fresher, more recent search results. *Google Inside Search*. Retrieved from https://search.googleblog.com/2011/11/giving-you-fresher-more-recent-search.html.
9. John Sutter. (2011, September 21). Santorum asks Google to clean up search results for his name. *CNN.com*. Retrieved from http://www.cnn.com/2011/09/21/tech/web/santorum-google-ranking/.
10. Alexander Burns. (2011, September 20). Santourm: Google spreads "filth". *Politico*. Retrieved from http://www.politico.com/story/2011/09/santorum-google-spreads-filth-063952#ixzz1YZ3jT247.
11. Google, 'Removal Policies'. Retrieved from https://support.google.com/websearch/answer/2744324?vid=1-635805400558656442-723298935.
12. David Segal. (2011, February 12). The dirty little secrets of search. *New York Times*. Retrieved from http://www.nytimes.com/2011/02/13/business/13search.html.
13. Matt Rosoff. (2011, February 23). Google busts Overstock.com for search spam. *Business Insider*. Retrieved from http://www.businessinsider.com/google-busts-overstock-for-search-spam-tactics-2011-2.
14. John Sutter. (2011, September 2011). Santorum asks Google to clean up search results for his name. *CNN.com*. Retrieved from http://www.cnn.com/2011/09/21/tech/web/santorum-google-ranking/.
15. Matthew Panzarino. (2012, December 12). Google tweaks image search algorithm and SafeSearch option to show less explicit content. *The Next Web*. Retrieved from http://thenextweb.com/google/2012/12/12/google-changes-search-image-safesearch-explicit/.
16. Laura Sydell. (2012, January 2012). How Rick Santorum's 'Google problem' endured. *NPR*. Retrieved from http://www.npr.org/2012/01/06/144801671/why-santorums-google-problem-remains.
17. Moz, 'Google algorithm change history'. Retrieved from http://www.seomoz.org/google-algorithm-change.
18. Brian Christian. (2012, April 25). The A/B test: Inside the technology that's changing the rules of business. *Wired*. Retrieved from http://www.wired.com/2012/04/ff_abtesting/.

19. Finding more high-quality sites in search. (2011, February 24). *Google Official Blog*. Retrieved from http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html.
20. Search quality highlights: 40 Changes for February. (2012, February 27). *Google Inside Search*. Retrieved from http://insidesearch.blogspot.com/2012/02/search-quality-highlights-40-changes.html.
21. Though Sullivan received this confirmation directly, Google did not issue this explanation publicly, nor did the company publicly discuss the change in Savage's site ranking. Settling questions about the specific implications of an algorithm tend to remain frustratingly murky and open to dispute, and not just because the algorithm itself is kept secret.
22. Miranda Miller. (2012, March 1). Spreading Santorum loses its frothy spot atop Google. Retrieved from http://searchenginewatch.com/article/2156606/Spreading-Santorum-Loses-Its-Frothy-Spot-Atop-Google.
23. Learn about Former Senator Rick Santorum. Retrieved from http://www.ricksantorum.com/about_rick.

Disclosure statement

I am currently employed by Microsoft, a competitor to Google in the area of search. My research for this essay was conducted before that employment began. Nevertheless, it could be perceived to be a conflict of interest. I hope it is clear that the aim of the essay is not to criticize Google's specific role in this incident, but to use the case to raise broader questions about algorithmic intermediaries and public discourse.

Notes on contributor

Tarleton Gillespie is a principal researcher at Microsoft Research, an adjunct associate professor in the Department of Communication and the Department of Information Science at Cornell University. He is co-founder of the blog *Culture Digitally*, the author of *Wired Shut: Copyright and the Shape of Digital Culture* (MIT, 2007) and co-editor of *Media Technologies: Essays on Communication, Materiality, and Society* (MIT, 2014) [email: tlg28@cornell.edu].

References

Ananny, M. (2015). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology & Human Values*, *41*(1), 93–117.

Barocas, S., Hood, S., & Ziewitz, M. (2013). Governing algorithms: A provocation piece. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2245322

Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York, NY: Portfolio.

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, *11*(6), 985–1002.

boyd, d., Levy, K., & Marwick, A. (2014). The networked nature of algorithmic discrimination. Open Technology Institute. Retrieved from http://www.danah.org/papers/2014/DataDiscrimination.pdf

Brunton, F. (2013). *Spam: A shadow history of the Internet*. Cambridge, MA: The MIT Press.

Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. Cambridge, MA: The MIT Press.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180.

Cheney-Lippold, J. (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, *28*(6), 164–181.

Citron, D., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209

Crawford, K. (2015). Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology & Human Values*, *41*(1), 77–92.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415.

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.

Gillespie, T. (2012). Can an algorithm be wrong? *Limn*, *1*(2). Retrieved from http://escholarship.org/uc/item/0jk9k4hj

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). Cambridge, MA: MIT Press.

Gillespie, T., & Seaver, N. (2015). Critical algorithm studies: A reading list. *Social Media Collective*. Retrieved from http://socialmediacollective.org/reading-lists/critical-algorithm-studies/

Gitlin, T. (1980). *The whole World is watching: Mass media in the making & unmaking of the new left*. Berkeley, CA: University of California Press.

Graham, S. D. N. (2005). Software-sorted geographies. *Progress in Human Geography*, *29*(5), 562–580.

Granka, L. A. (2010). The politics of search: A decade retrospective. *The Information Society*, *26*(5), 364–374.

Grimmelmann, J. (2008). The Google dilemma. *New York Law School Law Review*, *53*, 939–950.

Grimmelmann, J. (2014). Speech engines. *Minnesota Law Review*, *98*(3), 868–952.

Grosser, B. (2014). What do metrics want? How quantification prescribes social interaction on Facebook. *Computational Culture*, *4*. Retrieved from http://computationalculture.net/article/what-do-metrics-want

Hallinan, B., & Striphas, T. (2014). Recommended for you: The Netflix prize and the production of algorithmic culture. *New Media & Society*, *18*(1), 117–137.

Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society*, *16*(3), 169–185.

Kitchin, R. (2014). Thinking critically about and researching algorithms. Retrieved from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2515786

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, *15*(5), 769–787.

Marx, G. T. (2009). A tack in the shoe and taking off the shoe: Neutralization and counter-neutralization dynamics. *Surveillance & Society*, *6*(3), 294–306.

Naron, G. R. (1991). With Malice toward all: The political cartoon and the law of libel. *Columbia-VLA Journal of Law & the Arts*, *15*, 93–116.

Noble, S. U. (2012). Missed connections: What search engines say about women. *Bitch Magazine*, *54*, 36–41.

Pariser, E. (2012). *The filter bubble: How the new personalized web is changing what we read and how we think* (Reprint ed.). New York, NY: Penguin Books.

Pasquale, F. (2015). *The Black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Poon, M. (2007). Scorecards as devices for consumer credit: The case of Fair, Isaac & Company Incorporated. *The Sociological Review*, *55*(s2), 284–306.

Rieder, B. (2012). What is in pageRank? A historical and conceptual investigation of a recursive status index. *Computational Culture*, *2*. Retrieved from http://computationalculture.net/article/what_is_in_pagerank

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Paper presented to "Data and Discrimination," a pre-conference of the 64th annual meeting of the International Communication Association, Seattle, WA, USA.

Schudson, M. (1989). The sociology of news production. *Media, Culture & Society*, *11*(3), 263–282.
Seaver, N. (2013). Knowing algorithms. *Media in Transition 8*. Cambridge, MA. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/opre.42.2.201
Thompson, J. B. (2005). The new visibility. *Theory, Culture & Society*, *22*(6), 31–51.
Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency *Colorado Technology Law Journal*. v13 n2. Retrieved from http://ctlj.colorado.edu/wp-content/uploads/2015/08/Tufekci-final.pdf
Ziewitz, M. (2015). Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values*, *41* (1), 3–16.