fi®st m¤ñd@¥
PEER-REVIEWED JOURNAL ON THE INTERNET

Vernacular resistance to data collection and analysis: A political theory of obfuscation

by Finn Brunton and Helen Nissenbaum

Abstract

Computer–enabled data collection, aggregation, and mining dramatically change the nature of contemporary surveillance. Refusal is not a practical option, as data collection is an inherent condition of many essential societal transactions. We present one vernacular response to this regime of everyday surveillance, a tactic we call *obfuscation*. With a variety of possible motivations, actors engage in obfuscation by producing misleading, false, or ambiguous data with the intention of confusing an adversary or simply adding to the time or cost of separating bad data from good. Our paper develops a political philosophy of obfuscation, linking contemporary and historical cases to develop a descriptive account of obfuscation that is able to capture key commonalities in systems from radar chaff to BitTorrent.

Contents

---

1. Introduction: The problem of data gathering — Asymmetries of power and knowledge

What is obfuscation?

Supermarkets and grocery chains have always been in the data business, as well as the food business: with small profit margins and a product that can quickly spoil, they pay close attention to inventory, purchasing patterns, and geography. The introduction of store "loyalty cards" perfectly fit a decades–long pattern: rewarding loyal customers with additional discounts in return for better data, which could inform mailings, coupon campaigns, even which products to shelve together. So far, so normal — but the appearance of "loyalty cards," with their rather sinister Orwellian name, and direct connection of data collection with access to sales and discounts, sparked a strange revolt. Customers engaged in boycotts and tongue–in–cheek protests, but as loyalty cards became more common, and apparently permanent, strategies appeared to mitigate the perceived loss of privacy without entirely giving up the cards, and therefore the savings. Groups formed loyalty card swapping pools online, circulating the cards by mail or meetups; others created armies of clone shoppers by duplicating their cards over and over and distributing them to friends and strangers; households of roommates shared a single card. Whether because they resented the lack of choice — the way access to discounts effectively forced you to pay extra for shopping without a card — or worried about the unknown fate of their shopping data, customers found ways to make the data gathered about them less reliable, less useful, for its conjectured purposes. These defensive projects, the objections that sparked them, and the context in which they hoped for results constitute a form of vernacular resistance to data gathering and aggregation that we call *obfuscation*. By obfuscation, we mean producing misleading, false, or ambiguous data to make data gathering less reliable and therefore less valuable.

The fundamental anxiety the loyalty card protests speak to is obvious. Computer–enabled data collection, aggregation, and mining dramatically change the nature of contemporary surveillance. Innocuous traces of everyday life submitted to sophisticated analytics tools developed for commerce and governance can become the keys for stitching disparate databases together into unprecedented new wholes. This data is often gathered under conditions of profound power imbalance. Simply refusing to contribute to these profiles and collections is not a practical option: being profiled is the condition of many essential transactions, from connecting with friends in online social networks to shopping and traveling and engaging with public and private institutions.

In this paper, we develop a political philosophy of obfuscation. Linking contemporary and historical examples, we provide a descriptive account of obfuscation that captures key commonalities in systems ranging from chaff, which fills radar's sweep with potential targets, to the circulating exchanges of supermarket loyalty cards that muddle the record of purchases, to BitTorrent systems protecting their users from legal action by producing records of many IP addresses, only a few of which may be engaged in file sharing. Through these and other cases we can begin to clarify obfuscation among the other forms of resistance to surveillance, whether that surveillance takes the form of consumer data aggregation (supermarkets, or companies like Acxiom), monitoring for intellectual property violations (RIAA and MPAA), targeted advertising (sites like Google and Facebook), or police actions by repressive governments (which we will see addressed by obfuscation tactics within platforms for secure private conversation like Tor).

Additionally, we distinguish and evaluate different modes of obfuscation as well as motivations and power topologies of key actors: Are obfuscation tactics typically the response of the weak against the strong, adopted by those outside of circles of power and influence, or vice versa? Our political philosophy of obfuscation also addresses normative questions of legitimacy, asking whether smokescreens to avoid monitoring are morally defensible — ever, never, or sometimes? Under what conditions in the political landscape of surveillance are obfuscation's deceptive tactics acceptable? They can be deemed legitimate assertions of autonomy, or become problematic instances of economic free ridership (relying on others to be less conscientious in muddying their tracks and therefore better targets); they can be justified in resisting the obligation to acquiesce to monitoring, or be destructive acts, poisoning the wells of collective data. Obfuscation, as a tactic both personal and political, offers a platform for studying legitimate and problematic aspects of surveillance and resistance in an age of ubiquitous data capture.

Data gathering and its myths

Consider this scenario from Brad Templeton, chair of the Electronic Frontier Foundation, about "time–traveling robots from the future" (Templeton, 2009) [1]. These machines, with more powerful hardware and sophisticated software than we have today, will come back in time and subject us to total surveillance; they will connect the discrete — and, we thought, discreet — dots of our lives, turning the flow of our private lives into all–too–clear, all–too–human patterns, shining their powerful analytic light into the past's dark corners. These robots from the future are mercenaries, working for anyone wealthy enough to employ them: advertisers and industries, governments and interested parties. We are helpless to stop their work, as they collate and gather prior history, because we cannot change our past actions.

> The most mundane points of contact with contemporary life involve the involuntary production of data on our part: passing security cameras, withdrawing cash, making purchases with a card, making phone calls, using transit ...

In this science fiction story our contemporary situation is reflected. We are constantly generating data and this data is not going away. It is subject to increasingly powerful tools of aggregation and analysis over time — time–traveling robots from the future, which can discern things in the past that we never imagined would become visible. The most mundane points of contact with contemporary life involve the involuntary production of data on our part: passing security cameras, withdrawing cash, making purchases with a card, making phone calls, using transit (with a MetroCard or FasTrak tag, Oyster, Octopus, Suica, E–ZPass) — to say nothing of using the Internet, where every click and page may be logged and analyzed, explicitly providing data to the organizations on whose systems we interact. This data can be repackaged and sold, collected and sorted and acquired by a variety of means, and re–used for purposes of which we, the monitored, know nothing, much less endorse [2]. The importance and scale of this movement of collected personal data can be discerned from the size of the services industry that has sprung up around it, which includes companies like ChoicePoint (a specialist in collating data on individuals for security analysis, background checks and the like, purchased in 2008 for 4.1 billion dollars) and Acxiom, the world's largest processor of consumer data, with more than a

billion dollars in revenue for the 2010 fiscal year [3]. The unreliability of the businesses and public–private partnerships in this industry gives data mobility still more sinister dimensions, as materials are stolen, leaked, sold improperly or turned to very problematic ends by governments — ChoicePoint's sale of 145,000 records to identity thieves being one particularly egregious example [4]. The nature of these businesses, acquiring new data sets to add to their existing collections, points to the final area of concern. Multiple databases consolidated and cross–referenced, with incidental details linking previously disconnected bodies of information, produce a far more significant whole than any one part would suggest: identities, tendencies, groups and patterns with both historically revelatory and predictive power [5].

There is another side to the argument, however. Counteracting the visions of doom, "Big Brother," exposure, oppression, surveillance, and losses of privacy and freedom are the celebratory visions of enlightenment, knowledge, transparency, understanding, efficiency, and security through data analysis. For every image of an aggressively surveilled population, shamelessly manipulated by advertisers and coerced by government, there are models of a society where disasters are prevented and inefficiencies minimized, and deep data, available to the full range of analytics tools, enables objectively good decisions and resource allocation. You get what you want before you knew you wanted it, and crises are spotted early on and put in check with the skill of a grandmaster seeing future permutations of the board. These benefits, while potentially real, are often presented in a mythic argument, and we must understand the myths to ensure we can conscientiously analyze the benefits and trade–offs.

There are indeed some benefits we derive from the aggregation and analysis of data from individuals. Daily life in a society where the spaces of mobility and transaction are saturated with data–gathering devices and systems becomes a scientific instrument, quantifiable and open to study and optimization [6]. Vehicular sensors and in–car black boxes can improve gas mileage and turn every automobile accident into a learning experience for the prevention of future accidents [7]. Google's aggregation of search terms related to symptoms has enabled a kind of dashboard for flu outbreaks around the world, one that functions significantly faster than the Center for Disease Control's own reporting system [8]. The list goes on from fraud protection (the way your bank will contact you if atypical purchases are noted by the automated system) to enormous efficiencies and reduction of waste in the production and transportation supply chains that underlie contemporary businesses [9]. There are even benefits in areas as trivial as the uncanny Netflix movie recommendation engine! [10]

Readers undoubtedly will have noticed that the paragraph above, while accurate, is also a bit celebratory, with a strong rhetorical tilt, which we employ to emphasize mythic aspects of the benefits, the rhetoric of promise that activates and propels the advent of this technology in the public sphere. In this we are drawing on the work of Brian Pfaffenberger, who articulated the "technological drama": the deployment of technology as a major redistribution of power created by a new arrangement of technological artifacts and political values (Pfaffenberger, 1992). Central to this drama is the presentation of a myth, a grand vision of benefits constructed by a "design constituency," which draws on some of the "root paradigms" of the society, a move that invests the technology with the energy and momentum of a social movement for good and ill. (Pfaffenberger's own example is the role of nationalism and ethnic strife in the creation of the Sri Lankan irrigation system; similar examples could be made of "air–mindedness" and the cult of aviation in the first part of the twentieth century — in which avionics and the physical construction of planes were connected with potent national and cultural myths — and perhaps of the powerful narrative of sharing, love and community which is deployed around user–created content and open source software today. [11])

In the case of socio–technical systems of information aggregation and analysis, the "design constituency" — as Pfaffenberger terms "the groups and individuals who participated in the technology's design" — is a complex family of engineers, entrepreneurs and corporations with a shared interest in gathering, processing, and applying or commoditizing user data. Their task, along with building the technology itself, is to build a myth that the "impact constituency" — those who are disadvantaged or exploited by the new system (Neil Postman simply called them "losers" [12]) — cannot overcome. It must be mythic because, as Pfaffenberger points out, myths are far harder to argue with and better at motivating people than a necessarily vague list of estimated future benefits. Elements of the myth for the collection of personal data include the promise of quantification, instrumentalization and ever–greater efficiency, and the notion that we are undergoing a profound social shift in which personal privacy ceases to matter to all but society's malefactors. "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place," said Google's CEO Eric Schmidt; "At the end of the day, it's going to happen. Sites are going to fight it, but that data is going to become available," said Shion Deysarkar, CEO of information–gathering company 80legs; "That social norm [of information sharing] is just something that has evolved over time," said Facebook's founder Mark Zuckerberg [13].

There is a clear correlation between the evolution of a "social norm" — "only crooks care about not being constantly surveilled," "digital natives don't care about privacy" — and the bottom line of the companies making these claims. Root paradigms of "human nature," transparency, and the end of private life are invoked with the blatant obviousness of mythic statements for which one needs no real proof or argument. For our present purposes it suffices simply to point out the mythic dimension of the discussion, so we can keep it in mind while analyzing resistance to the project of aggregation and analysis. The power and the weakness of myth (and embedded in the one is always the source of the other, as with Achilles' heel) is that it profoundly simplifies, and the collection and use of individual data is a real, embedded, complex issue whose valuable and dangerous

consequences must be parsed away from both our fantasies and our fears of a more closely monitored and managed society.

Power and knowledge asymmetries

One fundamental problem with the systems of personal digital data collection and analysis is asymmetry, or rather, two asymmetries. First, the asymmetry of power: rarely do we get to choose whether or not we are monitored, what happens to information about us, and what happens to us because of this information. We have little or no say when monitoring takes place in inappropriate contexts, and is shared inappropriately with inappropriate others. Second, equally important, is an epistemic asymmetry: we are often not fully aware of the monitoring, and do not know what will become of the information produced by that monitoring, nor where it will go and what can be done to it — the time–traveling robots problem.

Your data is not accumulated in neutral circumstances, whether the collection involves surveillance at the level of infrastructure with which you must participate, or forms which have to be filled out to receive essential resources, or onerous terms of service to which you must consent before you can use an online product that has become vital to doing business. The context is often a major power imbalance, between individual consumers and major corporations, or citizens and governments. Obviously there is nothing inherently wrong with gathering data on individuals — it is lifeblood of the work of the epidemiologist, for example, and the starting point for many of the benefits mentioned above. It is in the combination of data gathering with authority and its interests where the problem begins.

It continues once our data has been collected. We don't know whether the company that gathers it will repackage and resell it, whether it will become part of the schedule of assets after a bankruptcy, or whether a private party like ChoicePoint will be collating it with public records and reassembling it in a very different context from our original provision. Data mining and related disciplines are complex and intellectually demanding; they often require resources of expertise, software and hardware that people outside large institutions do not possess. We don't have access to the other databases, nor the techniques and the training in mathematics and computer science, to comprehend what can be done with seemingly trivial details from our lives and activities, and how they can provide more powerful, total and revealing analyses than we could have anticipated [14]. The inconsequential and even benign can quickly become the problematic and sinister.

Furthermore, we don't know what future techniques and databases will enable — the time–traveling robots problem. Opportunities for the correlation of information tend increase with time. Institutions very rarely voluntarily destroy materials with as much potential as a rich database, and the mechanisms to extract value from databases are only going to get better. Materials from very different contexts, created in conditions of many different norms — telephone call logs, geolocative data, purchase records whether in person or online, demographic and personally identifying information, products of the data–generating machines that are social networking sites — can be combined, correlated and cross–referenced with less and less effort.

The lack of capacity to assess consequences in full is deeply troubling. We do not know all that they know about us, how they come to know it, or even who all the significant players might be. We cannot easily subject them to symmetrical analysis: such organizations might operate under the veil of national security or proprietary trade secrets, and we likely would not have the methods or the training to do anything with their data if we could get our hands on it. As people whose data is being collected, what we know of the situation is problematic, and what we do not know is substantial [15].

The intoxicating promise of data collection, aggregation and analysis, its mythic rhetoric of promise, has driven the complexity and breadth of its application, often to the dismay of critics who argue that they outstrip not only law and corporate best practice, but, indeed, our capacity to understand fully what happens and will happen to the data we produce beyond the first transaction to which we relinquish it. Which brings us to the question of obfuscation and its relationship to other, more formal means of redress for threats to individuals' privacy.

2. Standard means of redress

In theory, the ways out of our predicament of inescapable, ubiquitous, asymmetric collection and scrutiny of data are numerous and diverse, the palette of options familiar to anyone following the privacy debates. Each offers a prognosis for particular challenges, and each has shortcomings in relation to the asymmetries of data analysis.

If obfuscation is morally or politically problematic, why adopt it rather than relying on well–established mechanisms like user–opt–out, law, corporate best practice, and technology? We argue that each of these, while useful for certain types of threats, have not proven responsive to others, and all have particular short–term flaws, which could compound into a future that worries us. The first of these established — even reflexive

— approaches is the most common counterargument to the two asymmetries, the opt–out argument, which puts the responsibility on the shoulders of individuals whose data are being gathered. The other three are classic long–term, slow–incentive structures for creating social change; their gradual pace, and investment in existing interests, makes them problematic for near–term protection and sets the stage for self–directed and individually introduced strategies like obfuscation.

The steady rhetorical drumbeat in the discussion around data privacy is that refusal is a personal responsibility. If you're so offended by the way these companies collect and deploy your data, just don't use their services — *opt out*. No one is forcing you. To which we reply: yes and no. Many of these systems are not mandatory yet (government systems and various forms of insurance being just two exceptions), but the social and personal cost of refusal is already substantial, and growing. We pay by loss of utility, efficiency, connection with others in the system, capacity to fulfill work demands, and even merely being able to engage in many everyday transactions. Those who balk have few options: the degree to which data collection is an integral part of modern life, from electronic financial transactions to using the Internet, renders any attempt at an "opt–out" life onerous at best and ridiculous at worst: cash transactions, off–the–books employment, and pay phones, without plane travel, a vehicle, health insurance, or Internet access except under carefully managed conditions. Free e–mail, social networks, and search engines might strike one as more optional than necessary but to sacrifice them would jettison the significant swathe of social life mediated online. To rely entirely on personal choice is to leave all but the most dedicated and privacy obsessed at the mercy of the more conventional means of regulation — or resistance, with whatever means and capacities they have available.

The inability to live in the contemporary world without participating in data collection is only one element of the opt–out fallacy. Any real opt–out policy would also have to offer the granularity of the process of aggregation and analysis itself, allowing you to make choices that lay between the extremes of refusal and compliance. An opt–out of consequence would enable the receipt of certain benefits in return for a degree of use, data that could be gathered or deployed only in certain contexts or for certain purposes, for a set period of time, etc. This does not presently exist, and implementing it relies heavily on the diligence and good behavior of private corporations [16].

Which brings us to *corporate best practice*. Private sector efforts are hampered by the fact that companies, for good reasons and bad, are the major strategic beneficiaries of data mining. It is not too much to say that the contemporary consumer economy runs on data: Surveys, conversion and customer retention analysis, demography, targeted advertising, and data collected at the point of sale that feeds back through the entire supply chain, from the just–in–time production facility to the trendspotting system. Whether the company is in the business of gathering, bundling and selling individual data, like DoubleClick and ChoicePoint, or has relied on the data generated and provided by its customers to improve its operations, like Amazon and WalMart, or is based on user data–driven advertising revenue (Google's earnings reports reveal that 98 percent of its revenue is advertising), or subcontracts the analysis of consumer data for purposes of spotting credit, insurance, or rental risks, it is not in their interest to support general restraints on access to information.

Given the competitive disadvantage, any individual company going out on a limb risks losing the returns on customer, client, consumer, and even patient data. Further, the liquidity and portability of data renders any piecemeal strategy of relinquishment highly problematic, because material of little consequence in the context of one company can become part of a serious breach of privacy in another with access to a richer or better–managed database. (We are focusing primarily on the aggregation of online user data in this paper, but the possibility for violations of privacy though data collection and correlation is still larger in other industries, like the financial sector.) Finally, in a capitalist economy such as the United States, companies tend to resist any restriction at all on actions and practices that promise profit. Leaving it to the private sector to lead the way towards restraints on access to personal data, without at least some prodding, is like leaving it to the proverbial fox to guard the henhouse.

Law and regulation, historically, have been central bulwarks of personal privacy, from the Fourth Amendment of the U.S. Constitution to the E.U.'s data protection requirements and directives. While our laws will likely be the eventual site of conversation in which we answer, as a society, hard questions about the harvesting and stockpiling of personal information, it operates slowly, and whatever momentum propels them in the direction of protecting privacy in the public interest it is amply counterweighted by opposing forces of vested corporate and other institutional, including governmental, interests. In the meantime and in the near term, enormous quantities of personal data are already in circulation, packaged, sold, and provided freely and growing by the day.

Finally, there is great interest among the technical, particularly research, community in *engineering systems* that "preserve" and "enhance" privacy, be it in data mining, surfing or searching the Web, or transmitting confidential information. Detecting data provenance, properly anonymizing datasets, generating contextual awareness, and providing secure, confidential communication: mechanisms supporting these goals pose technical challenges, particularly when embedded in the real world or when working against the grain of features native to infrastructural systems such as the Web. Furthermore, no matter how convincing the technical developments and standards, adoption by key societal actors whose organizations and institutions mediate much data flow is another matter and fraught with politics.

Tools offered to individual directly, such as Tor and other proxy servers, are praiseworthy and valuable but the fact remains that they are not widely understood or deployed outside the relatively small circles of those who are already quite privacy–aware and technologically sophisticated. Additionally, there are utility costs: Tor can be slow, for example, and blocked by many large Web sites. All privacy–protecting technologies entail trade–offs, and those required by robust approaches like Tor have thus far kept their adoption relatively small.

We are not questioning the ability of law, the private sector, and technology to provide relief to individuals from unfettered monitoring, gathering, mining, and profiling, only that the wait for relief from these sources is likely to be long. The status quo offers too much gain from the power and epistemic asymmetries that define and entrench our predicament, and all these approaches still leave a gap. From our specific problem of the gathering and analysis of individual data we turn to an array of historical and contemporary examples of obfuscation so we can see it as a general strategy with many different forms, media, and motives. These examples illustrate some of the ways obfuscation has worked, and highlight systematic features that will be relevant to its evaluation, before we return to its application and related concerns for our particular moment and crisis.

3. Obfuscation in practice: Cases and examples

Obfuscation in its broadest and most general form offers a strategy for mitigating the impact of the cycle of monitoring, aggregation, analysis, and profiling, adding noise to an existing collection of data in order to make the collection more ambiguous, confusing, harder to use, and therefore less valuable. (We chose "obfuscation" for this purpose because of its connotations of confusion, ambiguity and unintelligibility, seeking to distinguish it from other strategies involving concealment or erasure, such as cryptography.) Obfuscation, like data gathering, is a manifold strategy carried out for a variety of purposes, with a variety of methods and perpetrators. Obfuscators may band together and enlist others, or produce misleading information on their own; they might selectively respond to requests for information, or respond so excessively that their contribution skews the outcome. They may engage in obfuscation out of a simple desire to defend themselves against perceived dangers of aggregation, in resentment of the obvious asymmetry of power and knowledge, to conceal legitimate activities or wrongdoing, or even in malice, to render the system of data collection as a whole worthless. This diversity of purposes, methods and perpetrators is reflected in the wide range of forms taken by obfuscation tactics.

These forms, across a range of media and circumstances, can be loosely clustered around four themes: relying on temporal limitations; requiring the "network effect" of cooperation or collaboration by groups of obfuscators; selectively interfering with data; and rendering data ambiguous and doubtful for the long term.

3.1. Time–based obfuscation

Whereas some forms of obfuscation try to inject doubt into the data permanently, time–based obfuscation, in many ways the simplest form of the practice, adds an onerous amount of processing in a situation where time is of the essence. *Chaff* offers a canonical example: The radar operator of the Second World War tracks a plane over Hamburg, guiding searchlights and anti–aircraft guns in relation to a phosphor dot whose position is updated with each sweep of the antenna. Abruptly the planes begin to multiply, their dots quickly swamping the display. The plane is in there somewhere, impossible to locate for the presence of all the "false echoes." The plane has released chaff, strips of black paper backed with aluminum foil and cut to half the target radar's wavelength, floating down through the air, thrown out by the pound and filling the system with signals. Chaff has exactly met the conditions of data the radar is configured to look for, and given it more planes, scattered all across the sky, than it can handle. Knowing discovery to be inevitable, chaff uses the time and bandwidth constraints of the discovery system against it by creating too many potential targets (in this regard, Fred Cohen terms it the "decoy strategy," and we can indeed consider obfuscation as the multiplication of plausible data decoys). (Cohen, n.d.) That the chaff only works briefly, as it flutters to the ground, and is not a permanent solution, is irrelevant under the circumstances; it only needs to work well enough for the time it will take the plane to get through.

Another contemporary example is the practice of *quote stuffing* in high frequency trading (HFT). (To be clear, quote stuffing is still only a theoretical obfuscation project, a plausible explanation for recent bursts of anomalous activity on the stock market.) The rarefied world of high–frequency trading (HFT) is built on algorithms that perform large volumes of trades far faster than humans, taking advantage of exceedingly minute spans of time and differences in price that would not be worth the attention of a human trader, if it were even physically possible for them to act on the change in price before the advantage was gone. Timing has always been critical to trading, but in HFT thousandths of a second separate profit and loss, and complex strategies to accelerate your trades and retard those of your competitors have resulted.

Analysts of market behavior began to notice unusual patterns of HFT activity over the summer months of 2010 — bursts of quote requests for a particular stock, sometimes thousands a second. Such activity seemed to have

no economic rationale, but one of the most interesting and plausible theories is that these bursts are an obfuscation tactic in action. "If you could generate a large number of quotes that your competitors have to process, but you can ignore since you generated them, you gain valuable processing time" (Nanex, 2010). Unimportant information, in the form of quotes, is used to crowd the field of salient activity, so the generator of the unimportant data can accurately assess what is happening while making it more difficult for their competitors to do so in time. They create a cloud of fog that only they can see through. None of these patterns would fool or even distract an analyst over a longer timescale — their artificial and insignificant character is obvious — but in the sub–split–second world of HFT, the act of having to observe and process this hiss of activity is enough to make all the difference.

As we will study further below, quote stuffing is an example of an obfuscation tactic that can actually be damaging, rather than merely annoying, to the system it uses. "This is an extremely disturbing development, because as more HFT systems start doing this, it is only a matter of time before quote–stuffing shuts down the entire market from congestion." (*Ibid.*) So much information can overload the physical infrastructure of the exchange like chaff dots swamping a radar display.

Finally, two examples of time–based obfuscation in thoroughly concrete contexts. The *"Craigslist robber"* offers a minor but illustrative example of obfuscation as a practice turned to criminal ends. At 11 AM on Tuesday, 30 September 2008, a man dressed like an exterminator in a blue shirt, goggles and a dust mask, and carrying a spray pump, approached an armored car parked outside a bank in Monroe, Washington, incapacitated the guard with pepper spray, and took a substantial amount of money. When the police arrived, they found 13 men in the area wearing blue shirts, goggles and dust masks — a uniform they were wearing on the instructions of a Craigslist ad which promised a good wage for maintenance work, which was to start at 11:15 AM at the bank's address. This is one of the few real–world examples of the recurrent trope of obfuscation in movies and television: the many identically dressed actors or objects confusing their pursuers as to the valuable one (Netter, 2008). Obviously it will only take a few minutes to determine that none of the day laborers is the bank robber — but a few minutes is all he needs.

Much of the pleasure and challenge of poker lies in learning to read people and deduce from their expressions, gestures, and body language whether they are bluffing, or pretending to hold a weaker hand in hopes of drawing a call. Central to the work of studying opponents is the "tell," some unconscious habit or tic an opponent will display in response to a strong or weak hand: sweating, a worried glance, leaning forward. Tells play such a crucial role in the informational economy of poker that players will use *false tells*, creating mannerisms which may appear to be part of a larger pattern [17]. According to common poker strategy, the use of a false tell is best reserved for a crucial moment in a tournament, lest the other players figure out that it is inaccurate and turn it against you in turn. A patient analysis of multiple games could separate the true from the false tells, but in the time–bound context of a high–stakes game the moment of falsehood can be highly effective [18].

3.2. Cooperative obfuscation

All of the cases described so far can be performed by a single actor (perhaps with some unwitting assistants), but other forms of obfuscation require the cooperation of others. They have the "network effect" of becoming more valuable as more people join. A powerful legend exemplifies this idea: the often re–told, factually inaccurate story that the king and population of Denmark wore the Yellow Star to make it impossible for the occupying Germans to distinguish and deport the Jews. While the Yellow Star was not used in Denmark for fear of arousing more anti–German feeling, "[t]here were documented cases of non–Jews wearing yellow stars to protest Nazi anti–Semitism in Belgium, France, the Netherlands, Poland, and even Germany itself" (Lund and Deak, 1990) [19]. The legend is a perfect story of cooperative obfuscation: a small group of non–Jews wearing the Yellow Star is an act of protest; a whole population, into which individual Jews can blend, is an act of obfuscation.

Loyalty card swapping pools provide a superb real–world example. Grocery stores have a long history of being in the technological vanguard when it comes to working with data — they have been early adopters of IBM computers, UPC and RFID identification systems, and the Electronic Data Interchange format to manage their logistics [20]. Early loyalty card programs were relatively innocuous, used to draw repeat customers, make extra margin from people who didn't use the card, and aid primitive data projects, such as, organizing direct mailings by zip code. The vast majority of grocers and chains outsourced the business of analyzing data to companies like ACNielsen and Catalina Marketing (Jackson, 2003).

Some worrying transfers of data happened nevertheless, changing context and value from innocuous to sinister. In 1999, the Robert Rivera case, a slip–and–fall in a Los Angeles Vons, led to a lawsuit in which the attorneys for the grocery chain threatened to disclose in court Rivera's history of alcohol purchases [21]. A string of similar cases over the years fed the already–Orwellian impression generated by the concept of "loyalty cards" in the popular imagination. Indeed, quite quickly after their widespread introduction came card–swapping networks, where people shared cards — initially in *ad hoc* physical meetings, and increasingly in large populations and over wide geographical regions enabled by mailing lists and online social networks — to obfuscate their data. Rob's Giant Bonus Card Swap Meet, for instance, started from the idea that a barcode

sharing system could enable customers of the D.C.–area supermarket chain to print out the barcodes of others, pasting them onto their cards [22]. A similar notion was adopted by the Ultimate Shopper project, mailing stickers of a Safeway loyalty card barcode and creating "an army of clones" accruing shopping data [23]. Cardexchange.org is devoted to exchanging cards by mail, presenting itself as a direct analogue to the physical meetups. These sites also act as clearinghouses for discussion, gathering notes, blog posts, news articles and essays on loyalty cards, debating the ethical implications of various approaches, and sharing theories and concerns. This is obfuscation as a group activity: the more who are willing to share their cards, the farther the cards travel, the more unreliable the data gets.

Another form of collective obfuscation appears in the argument for *participation in Tor*. Tor is a system designed to enable anonymous use of the Internet, through a combination of encryption and passing the message through many different independent "nodes." Imagine a message passed surreptitiously through a huge crowd to you. The message is a question without identifying information; as far as you know, it was written by the last person to hold it, the one who handed it to you. The reply you write and pass back vanishes into the crowd, following an unpredictable path. Somewhere in that group the writer receives his answer without you or anyone else knowing exactly who the writer was. That is to say: If you request a Web page while working through Tor, your request will not come from your IP address, but from an "exit node" (that last person who hands the message to its addressee) on the Tor system, along with the requests of many other Tor users. Data enters the Tor system and passes into a labyrinth of relays, computers on the Tor network (people in the crowd) that offer some of their bandwidth for handling Tor traffic from others, agreeing to pass messages sight unseen. The more relays there are, the faster the system is as a whole, and if you are already using Tor to protect your Internet traffic it is simple to turn your computer into a relay for the collective greater good. The Tor network — and the obfuscation of individuals on the network — improves as more people join in.

Obfuscation augments Tor's already considerable protective power, its designers point out. In return for running a Tor relay, as the FAQ says, "you do get better anonymity against some attacks. The simplest example is an attacker who owns a small number of Tor relays. He will see a connection from you, but he won't be able to know whether the connection originated at your computer or was relayed from somebody else." If someone has agents in the crowd of people — if they're running Tor relays for surveillance purposes — the agents can't read a message they pass, but they can notice who passed it to them. If you're on Tor and not running a relay, then they know you wrote the message you gave to them. But if you are letting your computer operate as a relay, the message might be yours or just one among many that you're passing on for other people. Did it start with you or not? The information is now ambiguous, and messages you've written are safe in a flock of other messages you pass along [24].

3.3. Selective obfuscation

All of the examples thus far have been about general methods of covering one's tracks. But what if you want this data to be useful without diminishing your privacy, or to interfere with some methods of data analysis but not others? This is the project of selective obfuscation. *FaceCloak*, for example, provides the initial steps towards an elegant and selective obfuscation–based solution to the problem of Facebook profiles (Luo, *et al.*, 2009). It takes the form of a Firefox plugin that acts as a mediating layer between a user's personal information and the social networking site. When you create a Facebook profile and fill in your personal information, including details such as where you live, went to school, likes and dislikes, and so on, FaceCloak offers you a choice: display this information openly, or keep it private? If you let it be displayed openly, it is passed to Facebook's servers like any other normal data, under their privacy policy. If you want to keep that data private, however, FaceCloak sends it to encrypted storage on a separate server only to be decrypted and displayed for friends you have authorized, when they browse your Facebook page (using the FaceCloak plugin.) Facebook never gains access to it. Furthermore, by generating fake information for the data that Facebook requires of its profiles, FaceCloak obfuscates its method — the fact that the real data lies elsewhere — from both Facebook and unauthorized viewers. As it passes your real data to the private server, FaceCloak generates a gender, with appropriate name, and age and passes those to Facebook. Under the cover of this generated, plausible non–person, you can connect and exchange with your friends, obfuscating the data for all others.

The theoretical goal for selective obfuscation has been outlined from a policy perspective as obfuscating the data for certain users or for the reconstruction of individual acts, in Gloria González Fuster's recommendations for EU data processing as *limiting the data to primary processing*: structuring the data such that it can be evaluated for its intended purpose, to which the data's subjects consent, but not for unanticipated analyses (González Fuster, 2009). In this scenario, data gathered for, say, a public health study would be suited to the process used for that study, difficult to use for other public health data mining, and impossible to reprocess for any other purpose.

Nam Pham and others on *privacy–preserving participatory sensing* shows us how this idea could work in practice, on an applied and mathematically sophisticated scale (Pham, *et al.*, 2010). Where a project like FaceCloak obfuscates the data for all but an authorized few, private participatory sensing obfuscates it beyond a certain degree of specificity — the data works generally, but not for identifying or tracking anyone in particular. Vehicular sensors, for instance, which can be used to create a shared pool of data from which to construct maps of traffic or pollution, raise obvious concerns over location–based tracking. However, Pham, *et al.* demonstrate

how to perturb the data, letting each vehicle continuously lie about its location and speed while maintaining an accurate picture of the aggregate.

3.4. Ambiguating obfuscation

Time–based obfuscation can be quickly seen through; cooperative obfuscation relies on the power of groups to muddy the tracks; selective obfuscation wishes to be clear for some and not others. Ambiguating obfuscation seeks to render an individual's data permanently dubious and untrustworthy as a subject of analysis. For example, consider the Firefox extension *TrackMeNot*, developed in 2006. Developed by Daniel Howe, Helen Nissenbaum, and Vincent Toubiana, TrackMeNot was designed to foil the profiling of users through their searches, a response to successive stories in the news — first the U.S. Department of Justice's request for Google's search logs, and later the surprising discovery by a *New York Times* reporter that some identities and profiles could be inferred even from anonymized search logs published by AOL [25]. Our search queries end up acting as lists of locations, names, interests, and problems, from which not only our identities can be determined, regardless of whether our IP addresses are included, but a pattern of our interests revealed. As with many of the previous cases of obfuscation, opting–out of Web search is not a viable choice for the vast majority of users. At least since 2006, search companies have acknowledged the problem of the collection of query logs, and have offered ways to address people's concerns, though they continue to collect and analyze these logs. In the meantime, the challenge remains to prevent any given stream of queries from being inappropriately revealing of a particular person's interests and activities.

TrackMeNot, therefore, automatically generates queries from a seed list of terms. These terms are initially culled from RSS feeds, and evolve over time, so that different users develop different seed lists. TrackMeNot submits queries in a manner that tries to mimic user search behaviors. This user may have searched for "good wi–fi cafe chelsea" but they have also searched for "savannah kennels," "freshly pressed juice miami," and "asian property firm," to say nothing of "exercise delays dementia" and "telescoping halogen light" — will the real searcher please stand up? The activity of individuals is masked by that of many ghosts, making the a pattern harder to discern, making it impossible to say, of any given query that it was the product of human intention rather than the automatic output of TrackMeNot.

Similarly, *BitTorrent Hydra* fights the surveillance efforts of anti–filesharing interests, by mixing genuine requests for bits of a file with dummy requests. The BitTorrent protocol breaks a file up into many small pieces, so that you can share those pieces, sending and receiving them simultaneously with other users. Rather than downloading an entire file from another user, as with the Napster model, you assemble the file's pieces from anyone else who has them, and anyone who needs a piece you have can get it from you. This many–pieces–from–many–people approach expedites sharing of files of all kinds, and quickly became the method of choice for moving large files, such as movies and music [26]. To help users of BitTorrent assemble the files they want, the system uses "torrent trackers," which log IP addresses that are sending and receiving files — if you're looking for these pieces of file *x*, users *a* through *n*, at the following addresses, have the pieces you need. Intellectual property groups, looking for violators, starting running their own trackers, which would serve the same function but gather the addresses so they could find major uploaders and downloaders of potentially copyrighted material. That network of users, swapping pieces of files, could be turned into a list of individuals responsible for piracy.

To protect these individuals, Hydra obfuscates by adding random IP addresses to the tracker, addresses that have been used for BitTorrent at some point. This means that periodically, as you request pieces of the file you want, you will be directed to another user that doesn't actually have what you're looking for. It is a small inefficiency for the BitTorrent system as a whole, but it makes address–gathering on the part of anti–piracy organizations much less useful. The tracker can no longer be sure that any one address was actually engaged in sharing that particular file. Doubt and uncertainty have been reintroduced to the system: can you sue with assurance? Rather than destroying the adversary's logs, or somehow concealing BitTorrent traffic, Hydra provides an "I am Spartacus" defense — recall the famous sequence in the Kubrick film, where one slave after another identifies themselves as the leader, and the one to be punished. Hydra does not avert data collection, but contaminates the results, making any specific case problematic and doubtful.

CacheCloak, meanwhile, has an approach to obfuscation suited to its domain of location–based services (LBSs) (Meyerowitz and Choudhury, 2009). LBSs take advantage of the locative technology in mobile devices to create various services, ranging from the trivial (FourSquare, which turns going places into a competitive game) to the lucrative (location–aware advertising) to the thoroughly useful (think of maps and nearest–object searches: "Where is the public restroom/movie theater/emergency room closest to me right now?"). The reader can see the classic rhetoric of balancing privacy against utility here, often presented to privacy's detriment, familiar from so many other cases in this field. If you want the value of an LBS, to be part of the network that your friends are on so you can meet if you are nearby, then you will have to sacrifice some privacy, and get used to the service provider knowing where you are.

CacheCloak offers a way through this seemingly intractable conflict. "Where other methods try to obscure the user's path by hiding parts of it," write the creators of CacheCloak, "we obscure the user's location by surrounding it with other users' paths" — the propagation of ambiguous data. In the standard model, your

phone sends your location to the service, and gets the information you requested in return. In the CacheCloak model, your phone predicts your possible paths and then fetches the results for several likely routes. As you move, you receive the benefits of locative awareness — access to what you are looking for, in the form of data cached in advance of potential requests — and an adversary is left with many possible paths, unable to distinguish the beginning from the end of a route, where you came from, and where you mean to go, still less where you are now. The salient data, the data we wish to keep to ourselves, is buried inside a space of other, equally likely data.

Finally, this form of obfuscation has been proposed as a defense against botnets. The technique of *botnet–resistant coding* operates on similar lines to quote stuffing. A botnet is a collection of malware–infected personal computers that can be controlled by a remote attacker, using system resources or snooping for data without their owners being aware of the activity. One of the more prolific of these botnets, known as Zeus, sits on the network looking for the patterns of data that suggest banking information; when found it sends the information — passwords, account details, and so on — back to its controllers. They will use it to make bank withdrawals or commit other forms of identity theft. The defensive solution proposed is an obfuscation move: very large quantities of completely plausible but incorrect information would be injected into the transactions between the user's computer and the bank (Rothschild and Greko, 2010). The bank knows how to filter the false information, because they generated it, but not the botnet does not. Faced with this source of confusion, attackers either move on to easier targets or waste resources trying to find the accurate needle in the bank's haystack.

---

4. The science of obfuscation

The examples we have compiled show something of the broad range of obfuscation practices, from foiling statistical analysis and escaping visual sensing to thwarting competitors in the stock market. Some methods take advantage of human biases, and others the constraints and loopholes of automated systems. Obfuscation is deployed for short-term misdirection, for legal deniability, to encourage an adversary to construct a flawed model of the world, and to change the cost–benefit ratio that justifies data collection. The swath of types, of methods, motives, means, and perpetrators are not surprising considering that obfuscation is a reactive strategy and, as such, a function of as many types of actions and practices as it is designed to defeat. Given this diversity, can a science of obfuscation exist? Can we create variables and parameters that will enable us to quantify its value and optimize its utility? Can we be sure obfuscation is working?

There are many variables, starting with the perpetrators of obfuscation. They may be lone individuals or groups of people working together; they may function privately, in an *ad hoc* manner, or systematically, in official capacities, under government authority. Obfuscators may be comparatively weak in relation to adversarial data gatherers, but it is possible for the strong to mislead the weak — a government deluging its citizens with data in useless or disingenuous acts of openness, or corporate malfeasance hiding behind hundreds of shell companies and a maze of paperwork. In the specific circumstances that have drawn us to consider the strategy of obfuscation, we have noted our concern for the power and epistemic asymmetries between information gatherers and holders, and information subjects. As such, our attention mainly focuses on configuration types and instances where the perpetrator of obfuscation is the weaker party.

In many cases, the knowledge asymmetry between the data gatherer and the obfuscator can lead to mistaken goals, which must be taken into account as well. Even as people were agitating against loyalty cards, far more serious work in mining individual data was happening in credit and associated areas of banking. In 2002, Canadian Tire's analysis of purchases on its credit cards revealed that people who buy premium motor oil, carbon monoxide detectors, birdseed, snow roof rakes and felt furniture pads are much better credit risks, and those who buy chrome skull accessories for their car or noise–generating exhaust systems are almost certain to default (Duhigg, 2009). American Express, as recent investigative journalism showed, applied analyses like these directly to risk management by lowering credit limits and changing APR in response to spending patterns (Lieber, 2009). These data mining projects were far less visible to users than loyalty cards, and therefore faced no obfuscating effort that we know of.

Helping to characterize different modes of obscuring data and how our definition of obfuscation fits in this larger picture, James Alexander and Jonathan Smith's theory of disinformation is useful (Alexander and Smith, 2010). In contrast with traditional information theory, after Claude Shannon's work, that assumes a sender and receiver who are eager to communicate equally motivated to their channel of unwanted noise, the scenarios Alexander and Smith consider involve reluctant senders, compelled to communicate and seeking ways to thwart, impede, and sabotage transmission: "[I]f we can't completely prevent the transmission, is there something we can do so that the message is received with some of its content missing? Or might we distort it enough that it is hard to recover or, better yet, it is entirely misleading?" [27]

A reluctant sender might opt for a disinformation attack that introduces noise into the transmission. Alexander and Smith identify two forms. In one, a "destructive disinformation attack" or "redaction," the sender damages an important subset of information, "reducing the content of the message, or, in information theoretic terms, …

increasing its entropy." [28] The problem with redaction is that the receiver will detect the loss of content and might, in turn, be able to thwart the sender with countermeasures of his own. Thus, Alexander and Smith prefer a "constructive disinformation" attack where a receiver is unaware the message has been tampered with. "What if, instead of just destroying the key information in the message, we were able to replace parts of it with false, but convincing, information? … Note that forging messages takes us outside the descriptive capability of conventional information theory: we are attempting to fool the receiver into believing that the communication system is behaving normally, when, in fact, it has failed." [29]

Obfuscation in the cases we discuss here addresses problems that are structurally similar to those Alexander and Smith tackle: when refusal to communicate or transmit information is not a realistic option for a sender who, nevertheless, is reluctant to do so unguardedly. Identifying points of overlap and difference between obfuscation and disinformation sheds light on both strategies. Although both disinformation and obfuscation obscure or degrade data streams, disinformation has a broader scope; while obfuscation covers only the addition of noise, disinformation may also involve tactics such as deletion or manipulation [30]. Yet obfuscation's scope is broader in embracing noise that is detectable by receivers and noise that is not; Alexander and Smith argue for the superiority of so-called constructive disinformation attacks.

The means and methods of obfuscation must be matched to specific scenarios in evaluating them: some call for stealthy obfuscation (or "positive disinformation"), while others require the opposite — the receivers need to know the data has been tampered with. In these instances, the goal may be to overwhelm their resources with false leads and red herrings, or raise the cost and difficulty of separating good from bad data until it becomes impractical to continue doing so, or to ensure that activities will fall in the shadow of reasonable doubt. A system like TrackMeNot may be able to achieve its goals whether or not it is known to be in use as long as some percentage of the queries it generates cannot be separated from actual user queries. Knowing that some part is suspect but not which, the collector has to either throw the lot out or take a loss of accuracy, and the protest against the collection has been lodged.

Before we ask of an obfuscation practice, "Does it work?," we have to ask, "What does it mean for it to work?" The goals or motives of those who obfuscate in relation to the goals and motives of the practice they wish to foil are as important to assessing success as the nature of means and methods adopted. The answer to "Does it work?" and the starting point of a science of obfuscation is "It depends on what you hope to achieve."

The goal may be modest, to decrease the value and immediate utility of data, to provide temporary respite from data analysis, to distract just long enough to complete a mission — a matter of minutes, in the case of chaff or the Craigslist robber with his identically–dressed clones, or microseconds in the case of quote stuffing. Other goals may need to affect data permanently, anticipating future, superior analytics techniques run on machines more powerful than those we have today. This long–term question challenges the effectiveness of TrackMeNot in the face of some as–yet–undeveloped form of semantic analysis can eventually separate out authentic queries. In some cases, the goal is to inject enough uncertainty to avoid blame (*e.g.*, incorrect addresses fed to the trackers by BitTorrent Hydra), and in others it is to wreak sufficient damage to make data unusable (*e.g.*, worthless information meant to make stolen banking data useless to botnet controllers). Some adversaries, put off by minor disutility, will abandon or limit use of compromised data in favor of easier, more fertile sources, while other adversaries will be dogged in pursuit, willing to spare no expense. Sometimes obfuscation relies on the compliance of others with the data gathering (the anti-botnet case is partially based on convincing the thieves to move to easier targets rather than working out how to mine the data), and sometimes on general insubordination: for the legendary adoption of the Yellow Star by gentiles in the legend of Denmark to actually be effective in protecting the Jewish population, the adoption must be general and widespread. To produce an analysis of how well the practice works and how to improve it requires an account of all these parameters, in the context of the obfuscator's goals and motives.

Can there ever be a science of obfuscation?

Can there ever be a science of obfuscation? With encryption, for example, algorithms have standard metrics based on objective measures such as key length, machine power, and length of time to inform community evaluations of their strength. By contrast, the success of obfuscation is a function of the goals and motives of both those who obfuscate and those to whom obfuscation is directed, the targets. We are tempted, for this reason, to characterize obfuscation as a relatively weak practice. Yet, when strong solutions, such as avoidance, disappearance, hiding (*e.g.*, through encryption) are not available and flat out refusal is not permitted, obfuscation may emerge as a plausible alternative, perhaps the only alternative. It simply has to be good enough, a provisional, *ad hoc* means to overcome the challenge that happens to be in its way. In our view, this contingency does not mean we throw up our hands to the challenge of a science. Although proof might not be achievable, it would nevertheless still be valuable to be able to assess how to optimize the value of various

obfuscation moves, even if only conditionally. Creating such a model is a challenge, to be sure. If there is to be a science of obfuscation it will need to identify key variables and create a systematic way of looking at the relationships between them. The set of variables will undoubtedly be hybrids of the social and the mathematical, including — goals (*i.e.*, time–based, ambiguating, selective), method (*i.e.*, whether group or individual, whether plausible data or obvious noise, whether hiding or protest), adversarial intent and resources (*i.e.*, time, opportunity cost), ratios (*i.e.*, of noise to signal), cost (*i.e.*, to obfuscator, to target), and more.

5. The politics of obfuscation

In the paper entitled "A tack in the shoe," Marx writes: "Criteria are needed which would permit us to speak of 'good' and 'bad,' or appropriate and inappropriate efforts to neutralize the collection of personal data." Along with the effectiveness of a particular practice, we must examine whether it is morally defensible. Now we ask the ethical and political questions that can be addressed to obfuscation, and balance the answers against monitoring, aggregation, mining, and profiling.

Given that obfuscation constitutes a counter–logic to data gathering and profile generation, an intervention to thwart it directly, we might conclude that obfuscation has no ethical or political valence of its own, only to the ends that it serves. If the surveillance in question is morally defensible, thwarting it by any means may be morally problematic, and, *mutatis mutandis*, obfuscation may be justified by unjust data practices. Prior to any analysis of ends, however, other moral and political considerations prompted by the very nature of obfuscation — wastefulness, dishonesty, free–riding, and more — deserve to be critically addressed.

Dishonesty

Implicit in obfuscation is an element of dishonesty — it is meant to mislead. Some people might balk at valorizing any practice that systematizes lying. (Some obfuscation approaches, such as that of CacheCloak, work around this problem by remaining ambiguous instead of providing untrue information — but such an approach depends on an informational relationship where queries can be left vague.) These critics might prefer encryption or silence to producing streams of lies. Whether lying, in general, can be morally justified is an exploration that clearly would take us too far afield from our subject, but that general discussion yields insights that are useful here. Excepting the Kantian who holds that lying is always absolutely wrong (famously, prescribing a truthful answer even to the murderer seeking one's friend's whereabouts), in many analyses there are conditions in which the proscription of lying may be relaxed. We must ask whether the general benefits of lying in a given instance outweigh costs, and whether valued ends are served better by the lie than truthful alternatives. There may be special circumstances in which lies may be excused, for example, if one is acting under duress, or lying to one party to keep a promise to another.

Free riding

Many forms of obfuscation rely on others' compliance. One protects one's own privacy by directing the adversary to targets who have not obfuscated their data, either because they do not have understanding or foresight, or are more trusting and accepting of data gathering. As the maxim of the wild has it, no need to be faster than the predator so long as one is faster than other prey. Obfuscation can be seen as two forms of free riding: taking advantage of the willingness of others to allow their data to be aggregated and processed, or enjoying the benefits of services while denying recompense to the targets of one's obfuscation — continuing to enjoy benefits without contributing to the cost by yielding one's own data into the pool. This is similar to a critique aimed at those who use adblocking software in the Web economy. Adblockers, plugins for Web browsers that prevent the advertisements on sites from displaying, threaten a major source of revenue that funds free content available online. Those who run adblockers can be among a privileged few to enjoy a quieter, faster–loading, ad–free Web, with free content underwritten by suckers who have not installed adblockers. So, too, many obfuscation strategies rely on the violation of privacy in general to protect the privacy of one who employs them. Loyalty card–swapping might also be understood in this light as participants enjoy the bounty of special offers while not contributing to the information pool that presumably enables these economies.

A key feature in the security of one system is the presence of other, more poorly secured systems; in the case of many adversaries, one simply needs to be slightly more secure to push the burden of exploits onto others (this is the model of the botnet–resistant coding project described above: to make the process of extracting the data sufficiently onerous that the thieves will find easier targets). Obfuscation, as a good–enough method, often leaves itself open to this critique, as many of its approaches rely on raising the cost of data gathering and analysis just enough to deter the surveillant, which relies on the cost generally being low.

Waste, pollution, and system damage

A common critique of obfuscation is that it wastes or pollutes informational resources — whether bandwidth and storage, or the common pools of data available for useful projects. In considering such accusations, we note

that "waste" is a charged word, implying that resources are used improperly, based presumably, on an agreed–upon standard. This standard could be challenged; what is wasteful according to one standard might be legitimate use according to another. How severe the disapprobation surely depends on the amount or degree of wastefulness, from virtually imperceptible to severe. However, noise introduced into an environment is not only wasteful but may taint the environment itself. This is particularly relevant for applications of data aggregation like supply chain efficiency, demographic analysis for medicine or government administration, or valuable scientific experiments. On a small scale, obfuscation may be insignificant — what can be the harm of marginal inaccuracy in a large database? On a large scale, however, it could render results questionable or even worthless. To take a recent case, the shopping logs of supermarket loyalty cards were used by the Centers for Disease Control and Prevention to identify a common purchase among a scattered group of people with *Salmonella*, trace that purchase to the source, and institute a recall and investigation, a socially valuable project which the widespread adoption of loyalty card swapping pools would have made much slower, or even, theoretically, impossible [31].

> If introducing noise into a system interferes with profiling, for example, it might harm the prospects of individuals, innocent bystanders, so to speak.

Data aggregation and mining is used not only to extract social utility but to guide decisions about individuals. If introducing noise into a system interferes with profiling, for example, it might harm the prospects of individuals, innocent bystanders, so to speak. FaceCloak demonstrates this problem: "[F]or some profile information (*e.g.*, an address or a phone number), it is ethically questionable to replace it with fake information that turns out to be the real information for somebody else." [32]. The risk is not only in the present, but holds for future uses not yet foreseen, the nightmare of the regularly incorrect United States No–Fly List writ large, or the mistakes of police profiling software compounded by a large pool of alternate, inaccurate names, addresses, activities, search terms, purchases, and locations. As a possible counterargument, however, if we believe that these databases and the uses to which they are put are malign, this bug becomes a feature. A database interlarded with ambiguously incorrect material becomes highly problematic to act on at all. As in the case of Tor relays example, or Hydra's anti–RIAA/MPAA tactic, the propagation of incorrect information might make the product of these systems inadmissible evidence for legal action.

Finally, waste includes the potential of damage, possibly fatal damage, to the systems affected by obfuscation. Consider quote stuffing in high–frequency trading, a move which, if broadly adopted, could actually overwhelm the physical infrastructure on which the stock exchanges rely with hundreds of thousands of useless quotes consuming the bandwidth. Any critique of obfuscation based in the threat of destruction must be specific as to the system under threat and to what degree it would be harmed.

Assessing the ethical arguments

The merits of each charge against obfuscation are not easily assessed in the abstract without filling in pertinent details. The overarching question that drives this paper is about obfuscation aimed at thwarting data monitoring, aggregation, analysis, and profiling, so we confine our evaluation to this arena, using cases we have introduced.

One consideration that is relevant across the board is ends. Legitimate ends are necessary, though, clearly, not always sufficient. Once we learn, for example, that the Craigslist robber used obfuscation to rob banks or that quote stuffing could bring down the Stock Exchange, it hardly seems relevant to inquire further whether the lies or free riding were justifiable. By contrast, banks seeking to foil botnets in order to protect customers' assets have an end worth pursuing. Establishing this point is no slam dunk, but it opens the way to further questions — whether the falsehoods, wastefulness, pollution, or free riding are justifiable. In several cases specifically aimed at foiling surveillance and profiling, the ends are contested. With BitTorrent Hydra, for example, it might be argued that the activities it is designed to hide are at the very least controversial and hence not able to sustain the legitimacy of the false leads. The question about ends need not be as straightforward as whether or not they are morally sound but whether or not they are proportional. The obfuscator running TrackMeNot need not have to show that Google's accumulating query logs is wrong outright; it may be enough to show that the accumulation of logs is disproportional to the legitimate ends. Similarly, an obfuscator might acknowledge the benefits implied by defenders of online behavioral targeting who ask, "What's wrong with relevant ads?" but believe that the degree of intrusiveness involved in tracking users across Web sites is disproportionate to them. The message the obfuscator is sending in such cases is not to cease but to bring objectionable practices in balance with the purported ends.

As the last point suggests, in cases such as TrackMeNot, CacheCloak, Tor relays, and loyalty card swapping, the arguments for and against obfuscation can become quite complex. Skeptics might agree, on the face of it, that the ends sought by these systems are legitimate, while still questioning the legitimacy of the methods used. To justify the falsehoods inherent in obfuscation, the ends must be unproblematic, and other aspects of the case taken into consideration — whether achieving the ends by means other than lying is viable, and what claim the targets of falsehood may have to "real" information. We must also consider broader contexts: when protection by law, technology, and corporate best practice fails, protection by obfuscation presents itself as the only resort. Further, these cases of individual versus Google, versus Verizon, versus government, etc. embody asymmetries of power and knowledge. Under duress and with little assurance that those extracting information can be trusted, the obligation to speak the truth is certainly lessened. Contrast this with highly controlled environments, such as a courtroom, where a myriad other constraints circumscribe the actions of all parties; we may still speak under duress but epistemic asymmetries are mitigated because of the strictures of context.

While deception may be justified by asymmetries and the absence of alternatives, other critiques remain. Wastefulness is a charge that may be leveled against systems such as TrackMeNot that "waste" bandwidth by increasing network traffic and "waste" server capacity by burdening it with search queries that are not, in reality, of interest to users. A cost–benefit or utilitarian assessment directs us to consider the practical question of how severe the resource usage is. Does the noise significantly, or even perceptibly undermine performance? In the case of search queries, which are short text strings, the impact is vanishingly small compared with the Internet's everyday uses at this point, such as video distribution, online gaming, and music streaming.

Additionally, it is not sufficient to hang the full weight of the evaluation on degree of usage — it is necessary to confront normative assumptions explicitly. There is irony in deeming video streaming a use of network but a TrackMeNot initiated search query a waste of network, or a TrackMeNot initiated query a waste of server resource but a user generated search for porn a use. This makes sense, however, once we acknowledge that the difference between waste and use is normative; waste is use of a type that runs counter to a normative standard of desired, approved, or acceptable use. The rhetoric of waste, however, begs to be scrutinized because while it may be dressed up as an objective, definable concept, in many cases it is speakers who inject and project their perspectives or interests into defining a particular activity as wasteful. To the extent that the Internet is a common or social resource, it is reasonable to expect that judging uses as legitimate and not wasteful be based on common and not parochial standards. Of course, these will be highly contested, but at least we will not be pre–empting these questions entirely. Those who turn to obfuscation must, therefore, be ready to defend their choices by referring to these standards. In the case of TrackMeNot, for example, much will depend on the success of the argument for privacy rights in search queries, in turn legitimizing the effort to ensure those rights one way or another [33].

The use/waste conundrum gains traction from another assumption about the information flows between individuals and the agencies and service providers that monitor and profile them. Individuals using FaceCloak or CacheCloak, even if not "wasting" resources according to widely held social standards, may still draw the ire of Facebook or location–based services. As businesses see it, the users in question are "wasting" their resources because they are depriving them of the positive externalities of personal information flows, which normally would enrich either their own data stockpiles or those of others to whom this data is sold or exchanged. Do we have reason to believe that the services in question are morally entitled to this positive externality — and, if so, could they be exceeding their entitlement by how the capture those flows or the uses they put them to (the problem of proportionality)? The difference between them and, say, mobile phone companies, is that the arrangement is a voluntary one. Thus, Facebook and Foursquare may claim there is an implied contract, or even one that is explicated in terms of service or privacy policy. If you use our service, you are bound by these terms; if you use the service differently, you are not only violating our terms but are free riding on our investment and the contributions of others. In other words, the obfuscator is a bad actor for violating an implicit agreement (if not contract) with the service providers and, moreover, benefiting from the fact that others are not. The problem of free riding on the contributions of others casts obfuscation efforts in an unseemly light. The obfuscator is presented as not so much the rebel as the sneak.

We hold off responding to these charges until we have discussed the problem of data "pollution" and the propagation of error and inaccuracy. These may be the trickiest of all, and get to the heart of obfuscation. The intention behind inserting noise into the data stream is precisely to taint the resulting body. But there are various ways it can be tainted and some may be more problematic than others. One misspelled name does not a ruined database make; at what point does inaccurate, confusing and ambiguous data render a given project or business effectively worthless? Obfuscation that does not interfere with a system's primary functioning but affects only secondary uses of information might be quite fair [34]. Further, while some obfuscation practices might confuse efforts to profile individuals accurately, they may not render aggregate analysis useless, for example, as in the case of Abdelhazer's work on perturbing individual data while retaining a reliable total picture. But what if none of these mitigations are possible? Where does this leave the ethics and politics of obfuscation?

It is clear that certain elements of free riding and data inaccuracy remain intractable, locking the obfuscator and data aggregator in a stalemate of disagreement. Those who engage in obfuscation and are prepared to benefit from the willingness of others to be monitored and profiled may believe that each group is merely acting

according to preference. Well and good, perhaps. The actions of those who choose obfuscation because they believe there are threats, harms, and violations of rights in respective data practices may be morally problematic. One answer to detractors who point this out might be "but they are no worse off than if I had not chosen to obfuscate." Another answer challenges detractors to hold data gatherers morally responsible for the wrongdoing. "Don't blame me for being fleet footed; it is the predator who is responsible for the demise of slower victims." A similar rebuttal to complaints over data inaccuracy directs blame to the data gatherers.

Those coerced into providing information into the data pool with insufficient assurance over how it will be used, where it will travel, how it will be secured, are being asked to write a blank check with little reason to trust the check's recipients. Under coercion, obfuscation is not a luxury but an action of last resort. When pushed to the corner, in cases where the issues of extra load on resources, free riding, and data tainting cannot be denied, where the obfuscator acts earnestly to resist the machinations of monitoring and analysis, obfuscation must be evaluated as an act of reasonable and morally sound disobedience.

6. Conclusions

Obfuscation, as we have presented it here, is at once richer and less rigorous than academically well–established methods of digital privacy protection, like encryption. It is far more *ad hoc* and contextual, without the quantifiable protection of cryptographic methods — a "weapon of the weak" to take a phrase from James Scott for the modes of resistance available to those at the wrong end of the asymmetries we have described. It is often haphazard and piecemeal, creating only a temporary window of liberty or a certain amount of reasonable doubt. And it is for precisely those reasons that we think it is a valuable and rewarding subject for study. The concept can be easily understood and inventively deployed, and lets us lower the stakes of resistance, making it possible for people coerced into compliance by necessity, circumstance or demand to push back.

Politically, as long as the ends are sound and we take care to avoid certain methods, obfuscation can be a force for good in our contemporary culture of data. These moves are a valuable resource in the defense of our privacy and freedom of action. We have provided an outline of the family, a number of examples, the parameters for quantification and improvement, and a view of the political and ethical problems and exigencies it creates. Now, we hope the community of privacy researchers and activists will help expand this idea. We face a number of further questions, beginning with one scientific, one moral, and one technical:

- Is it possible to create a meaningfully quantified science of obfuscation? Can we optimize different obfuscation tactics for different scenarios, and find weak points in the overall strategy?

- Does our description of obfuscation as viable and reasonable method of last–ditch privacy protection lead to the same political problems created by other systems of privacy preserving technology and possibilities like opt out — that is, putting the responsibility back on the private user and side–stepping the need to create a mature civil society around managing data?

- Are there methods for counter–profiling — figuring out how the profilers work to fine–tune our data strategies to best stymie them — that could be incorporated into the project of refining obfuscation?

Under duress, in the face of asymmetry, innovative methods for drawing the contextual lines of information flow will emerge; people will create models of informational security and freedom from invasive analysis, whatever claims profit–seeking CEOs make about "human nature" and its transformations. Obfuscation is often cheap, simple, crude, clever rather than intelligent, and lacks the polish or freedom from moral compromises that characterizes more total privacy solutions. Nonetheless it offers the possibility of cover from the scrutiny of third parties and data miners for those without other alternatives. It is the possibility of refuge when other means fail, and we are obliged both to document it, and to figure out if it can be made stronger, a more effective bulwark for those in need. 🇫🇲

About the authors

Finn Brunton is a postdoctoral researcher at New York University (NYU) in the Department of Media, Culture, and Communication, focusing on technological adaptation and misuse, and has written on topics including anonymity and WikiLeaks; he is currently preparing a book, *The Spew: A History of Spam*. He will be joining the faculty of the School of Information at the University of Michigan in the fall of 2011.
E–mail: finnbr [at] gmail [dot] com

Helen Nissenbaum writes on ethics, values and privacy in information technologies, design, and new media, and her work in these areas has been supported by the National Science Foundation, the Air Force Office of

Scientific Research, and other institutions. Her most recent book is *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford, Calif.: Stanford Law Books, 2010). She is a Professor of Media, Culture, and Communication & Computer Science and a Senior Faculty Fellow of the Information Law Institute at NYU. E–mail: hfn1 [at] nyu [dot] edu

Notes

1. The relevant passage: "The time traveling robots from the future that I am talking about are all of the people in this room who are working on AI … You are going to get better at face recognition, speech recognition, identifying people from their voices and so on. Those AIs from the future are going to be able to come into the past — not literally … — but metaphorically in that they will be able to search all of these databases that we build now with better tools. They will be able to look at all the video that is being recorded today and all the ATM machines you used and say, 'Where was Brad on February 7 of 2009? Oh, our modern face recognition software can look through those old records and find out.' The sins of the past will be visited upon you in the future with tools that you did not know existed."

2. For a cogent explanation of the additional processing to which gathered data can be subject, see González Fuster (2009).

3. A brief press release documenting the purchase of ChoicePoint by Reed Elsevier is available at http://www.reed-elsevier.com/mediacentre/pressreleases/2008/Pages/AcquisitionofChoicePointIncCompleted.aspx. For a journalistic account of the incredible scale of Acxiom's operation, see the CNN story on the company: http://money.cnn.com/magazines/fortune/fortune_archive/2004/02/23/362182/index.htm. For their fiscal year 2010 results, see http://www.acxiom.com/news/press_releases/2010/Pages/AcxiomAnnouncesFourthQuarterandFiscalYear2010Results.aspx. All accessed 12 December 2010.

4. The sale is well documented by the account in CSOonline, http://www.csoonline.com/article/220340/the-five-most-shocking-things-about-the-choicepoint-data-security-breach, and the reactions by the FTC and ChoicePoint have been collected in the Privacy Rights Clearinghouse "Chronology of Data Breaches" (see under 15 February 2005): http://www.privacyrights.org/ar/CPResponse.htm. This incident led to the thought–provoking "Model regime of privacy protection" proposed by Daniel Solove and Chris Jay Hoofnagle; see Solove and Hoofnagle (2005).

5. In making this argument we are drawing on our descriptions of this problem with reference to the received notion of privacy in Nissenbaum (1998; 1999).

6. In the area of medicine, for example, see Stead and Lin (2009).

7. For an excellent overview of the technical and legal dimensions of "event data recorders," or black boxes for use after car crashes, see http://www.harristechnical.com/cdr.htm (accessed 25 November 2010).

8. For the dashboard itself, see http://www.google.org/flutrends/. For a story analyzing the program, see http://www.guardian.co.uk/technology/2008/nov/12/google-health (accessed 11 October 2010).

9. See Subramani (2004), which provides both a breakdown of the ways data analysis in supply chain management systems benefits buyers, and also makes a cogent argument for its benefits to suppliers as well.

10. For a brief look at the Netflix Prize and the surprisingly deep problems posed by creating recommendation engines, see http://www.technologyreview.com/computing/23635.

11. On air–mindedness, the authoritative account is available from Wohl (1994; 2005).

12. Postman (1990): "Technological change, in other words, always results in winners and losers."

13. For Eric Schmidt's remark, see his interview with CNBC on 3 December 2009, as excerpted in http://www.youtube.com/watch?v=A6e7wfDHzew. For Deysarkar's, see his interview on ReadWriteWeb, 19 April 2010: http://www.readwriteweb.com/archives/bulk_social_data_80legs.php. For Zuckerberg's, see his interview with Mike Arrington, 8 January 2010, as video at http://www.ustream.tv/recorded/3848950; for a transcript of the relevant remarks, see http://www.readwriteweb.com/archives/facebooks_zuckerberg_says_the_age_of_privacy_is_ov.php (all accessed 16 December 2010).

14. See, for example, the description of these problems in Solove (2008) and Reiman (1995).

15. As one among many possible examples of our ignorance of the future uses to which our data may be put — whether it's records sold by an unscrupulous employee or left in a cab on a USB drive — see the business of scraping social network sites for their data, which can be bundled, sold and used without our ever being aware or giving consent to this use: http://www.readwriteweb.com/archives/bulk_social_data_80legs.php. For analysis of this situation from a specifically legal perspective, see Hildebrandt (2008) and Zarsky (2006).

16. See Barocas and Nissenbaum (2009) for an instance of this problem of consenting to data use after the fact.

17. An anecdotal account of false tells from poker player Phil Hellmuth, from Navarro (2006), can be found online at http://southerngaming.com/?p=62.

18. It's interesting to imagine a poker strategy based around more extensive use of obfuscation — a player generating a constant stream of mannerisms and typical tells, so that anything involuntary is difficult to parse out — but it would probably be so irritating as to get a player ejected!

19. To be clear, that the specific case of the Danes and the Yellow Star is fictional in no way detracts from their heroic wartime history of helping Jews hide and escape.

20. In a striking early example, a 1951 address Arthur C. Nielsen, of the Nielsen Ratings system, delivered to the Grocery Manufacturers of America included this offer: "The [special retail census] material is available either in the form of neat tabulations or on IBM tabulating cards — and in the latter case we are prepared to punch in the codes for your own sales territories or other data designed to make the material even more useful to you." Nielsen (1952).

21. *Privacy Journal*, March 1999, p. 5.

22. See Rob Carlson's site: http://epistolary.org/rob/bonuscard/, accessed 25 October 2010.

23. The Ultimate Shopper project: http://www.cockeyed.com/pranks/safeway/ultimate_shopper.html (accessed 19 October 2010).

24. As the FAQ points out, as a practical matter this may not make a difference to a truly empowered adversary with complete oversight of the traffic moving onto and off of your relay — a person who has agents on all sides of you, and knows what's been passed and what hasn't.

25. For the AOL search logs event, see http://www.nytimes.com/2006/08/09/technology/09aol.html; for the U.S. Department of Justice's Google request, see the original subpoena, http://www.google.com/press/images/subpoena_20060317.pdf, and the consequent ruling: http://www.google.com/press/images/ruling_20060317.pdf (accessed 15 August 2010).

26. See the ipoque ISP traffic analysis 2008/2009: http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009 (accessed 5 September 2010).

27. Alexander and Smith, 2010, p. 1.

28. *Ibid.*, p. 5.

29. *Ibid.*, p. 6.

30. Alexander and Smith (2010) give the example of switching the directional "N" in a message to a "W," creating the minimum number of detectable artifacts of tampering.

31. See, among many other stories, this summary of the *Salmonella* outbreak from *Business Week*: http://www.businessweek.com/ap/financialnews/D9EC5QUG0.htm.

32. Wanying Luo, Qi Xie, and Urs Hengartner, 2009. "FaceCloak: An architecture for user privacy on social networking sites," at http://www.cs.uwaterloo.ca/~uhengart/publications/passat09.pdf, p. 6.

33. Howe and Nissenbaum (2009). A related critique, raised by many of our discussants on the subject of waste

and pollution in obfuscation, is ecological: what is the carbon footprint of generating enough false information to protect privacy? Obviously, this depends on the specific tactic, but let us take the case of TrackMeNot and Google. We can make a simple calculation (the energy consumption of a server divided by the time it takes it answer a single query) and say that's the cost of a TMN query, but that's not actually accurate: the server consumes energy whether it's answering requests or not. The real question is whether, given sufficient adoption of TMN or a similar technology, Google would set up more server farms to ensure capacity with all the additional use, which would be a matter of significant environmental impact — as well as offering a significant possibility for social protest: either add more electricity–consuming servers … or stop mining our query data, in which case we will stop using TMN and your traffic will go back down to reasonable levels. As this example suggests, the question of obfuscation and the environment is an ethically complex one, and warrants further thought.

34. Again, see the analysis in González Fuster (2009), which provides a cogent explanation of and argument for the process of making data fit for an intended, "primary" use and unfit for further "secondary" — and unconsensual — uses.

References

James M. Alexander and Jonathan M. Smith. 2010. "Disinformation: A taxonomy," University of Pennsylvania Department of Computer and Information Science, Technical Report number MS–CIS–10–13, and at http://repository.upenn.edu/cis_reports/920/, accessed 25 April 2011.

Solon Barocas and Helen Nissenbaum, 2009. "On notice: The trouble with Notice and Consent," *Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information* (October), at http://www.nyu.edu/projects/nissenbaum/papers/ED_SII_On_Notice.pdf, accessed 25 April 2011.

Fred Cohen, n.d. "The use of deception techniques: Honeypots and decoys," *Fred Cohen & Associates*, at http://all.net/journal/deception/Deception_Techniques_.pdf, accessed 12 December 2010.

Charles Duhigg, 2009. "What does your credit–card company know about you?" *New York Times* (12 May), at http://www.nytimes.com/2009/05/17/magazine/17credit-t.html, accessed 25 April 2011.

Gloria González Fuster, 2009. "Inaccuracy as a privacy–enhancing tool," *Ethics and Information Technology*, volume 12, number 1, pp. 87–95.http://dx.doi.org/10.1007/s10676-009-9212-z

Mireille Hildebrandt, 2008. "Profiling and the rule of law," *Identity in the Information Society (IDIS)*, volume 1, number 1, pp. 55–70.http://dx.doi.org/10.1007/s12394-008-0003-1

Daniel Howe and Helen Nissenbaum, 2009. "TrackMeNot: Resisting surveillance in Web search," In: Ian Kerr, Valerie Steeves, and Carole Lucock (editors). *Lessons from the identity trail: Anonymity, privacy, and identity in a networked society*. Oxford: Oxford University Press, pp. 417–440.

Joab Jackson, 2003. "Cards games: Should buyers beware of how supermarkets use 'loyalty cards' to collect personal data?" *Baltimore City Paper* (1 October), and at http://www.joabj.com/CityPaper/031001ShoppingCards.html, accessed 25 April 2011.

Ron Lieber, 2009. "American Express kept a (very) watchful eye on charges," *New York Times* (30 January), at http://www.nytimes.com/2009/01/31/your-money/credit-and-debit-cards/31money.html, accessed 25 April 2011.

Jens Lund; reply by István Deák, 1990. "The legend of King Christian: An exchange," *New York Review of Books* (29 March), at http://www.nybooks.com/articles/archives/1990/mar/29/the-legend-of-king-christian-an-exchange/, accessed 25 April 2011.

Wanying Luo, Qi Xie, and Urs Hengartner, 2009. "FaceCloak: An architecture for user privacy on social networking sites," Proceedings PASSAT '09: 2009 IEEE International Conference on Privacy, Security, Risk and Trust (Vancouver, B.C.), pp. 26–33, and at http://www.cs.uwaterloo.ca/~uhengart/publications/passat09.pdf.

Gary T. Marx, 2003. "A tack in the shoe: Neutralizing and resisting the new surveillance," *Journal of Social Issues*, volume 59, number 2, pp. 369–390.http://dx.doi.org/10.1111/1540-4560.00069

Joseph Meyerowitz and Romit Roy Choudhury, 2009. "Hiding stars with fireworks: Location privacy through camouflage," *MobiCom '09* (Beijing, China), and at http://synrg.ee.duke.edu/papers/cachecloak.pdf, accessed 25 April 2011.

Nanex, LLC, 2010. "Analysis of the 'Flash Crash' Date of Event: 20100506, Part 4, Quote Stuffing" (18 June), at http://www.nanex.net/20100506/FlashCrashAnalysis_Part4-1.html, accessed 26 November 2010.

Joe Navarro, 2006. *Phil Hellmuth presents read 'em and reap: A career FBI agent's guide to decoding poker tells*. New York: Collins.

Sarah Netter, 2008. "Wash. man pulls off robbery using Craigslist, pepper spray," *ABC News* (1 October), at http://abcnews.go.com/US/story?id=5930862, accessed 25 April 2011.

Arthur Nielsen, 1952. "What's new in food marketing and marketing research: An address to Grocery Manufacturers of America at Hotel Waldorf–Astoria, New York, N.Y., November 12, 1951," New York: A.C. Nielsen Co.

Helen Nissenbaum, 1999. "The meaning of anonymity in an information age," *The Information Society*, volume 15, number 2, pp. 141–144, and at http://www.indiana.edu/~tisj/readers/abstracts/15/15-2%20Nissenbaum.html, accessed 25 April 2011; reprinted in Richard A. Spinello and Herman T. Tavani (editors). *Readings in cyberethics* boston: Jones and Bartlett, 2001.

Helen Nissenbaum, 1998. "Toward an approach to privacy in public: The challenges of information technology," *Ethics and Behavior*, volume 7, number 3, pp. 207–219; reprinted in Richard A. Spinello and Herman T. Tavani (editors). *Readings in cyberethics* boston: Jones and Bartlett, 2001.

Brian Pfaffenberger, 1992. "Technological dramas," *Science, Technology, & Human Values*, volume 17, number 3, pp. 282–312.http://dx.doi.org/10.1177/016224399201700302

Nam Pham, Raghu K. Ganti, Yusuf S. Uddin, Suman Nath, and Tarek F. Abdelzaher, 2010. "Privacy–preserving reconstruction of multidimensional data maps in vehicular participatory sensing," *Proceedings of WSN '2010: Seventh European Conference on Wireless Sensor Networks*, pp.114–130, and at http://research.microsoft.com/pubs/115793/privacy-EWSN10.pdf, accessed 25 April 2011.

Neil Postman, 1990. "Informing ourselves to death," speech given at the German Informatics Society (Stuttgart, 11 October); transcript at http://w2.eff.org/Net_culture/Criticisms/informing_ourselves_to_death.paper, accessed 24 November 2010.

Jeffrey Reiman, 1995. "Driving to the panopticon: A philosophical exploration of the risks to privacy posed by the highway technology of the future," *Santa Clara Computer and High Technology Law Review*, volume 11, number 1, pp. 27–44.

Fabian Rothschild and Peter Greko, 2010. "Botnet resistant coding: Protecting your users from script kiddies." paper presented at The Next HOPE (16 July), at http://thenexthope.org/talks-list/, accessed 15 October 2010.

Daniel J. Solove, 2008. "Data mining and the security–liberty debate," *University of Chicago Law Review*, volume 74, number 1, p. 343–362.

Daniel J. Solove and Chris Jay Hoofnagle, 2005. "A model regime of privacy protection," (Version 2.0; 5 April), GWU Law School Public Law Research Paper, number 132; GWU Legal Studies Research Paper, number 132, at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=699701, accessed 13 November 2010.

William W. Stead and Herbert S. Lin (editors), 2009. *Computational technology for effective health care: Immediate steps and strategic directions*. Committee on Engaging the Computer Science Research Community in Health Care Informatics, National Research Council of the National Academies. Washington, D.C.: National Academies Press.

Mani Subramani, 2004. "How do suppliers benefit from information technology use in supply chain relationships?" *MIS Quarterly*, volume 28, number 1, pp. 45–73.

Brad Templeton, 2009. "The evils of cloud computing: Data portability and single sign on," 2009 *BIL Conference* (Long Beach, Calif.); video at http://www.vimeo.com/3946928, accessed 5 October 2010.

Robert Wohl. 2005. *The spectacle of flight: Aviation and the Western imagination, 1920–1950*. New Haven, Conn.: Yale University Press.

Robert Wohl, 1994. *A passion for wings: Aviation and the Western imagination, 1908–1918*. New Haven, Conn.: Yale University Press.

Tal Z. Zarsky, 2006. "Online privacy, tailoring and persuasion," In: Katherine J. Strandburg and Daniela Stan Raicu (editors). *Privacy and technologies of identity: A cross–disciplinary conversation*. New York: Springer Science+Business Media, pp. 209–224.

Editorial history