DATA MINING, DOG SNIFFS, AND THE FOURTH AMENDMENT

An algorithm, designed to probe a database containing all personal data available to the government, sees that you have recently bought some fertilizer and a large truck, and that you have emailed someone with a .lb (Lebanon) email address. Seeing this red flag pop up on his computer, a government agent pulls your bank records, your Amazon and iTunes purchases, the domain names that you've recently visited, and a list of everyone you have recently phoned or emailed. Inspecting all of these records, the agent determines that you were merely asking your Lebanese uncle for advice on expanding your farm and makes a notation in his database. (He is also able to determine your religious affiliation, that you have an affinity for Steven Seagal movies, and that you have been having an affair, but is less interested in all of that.)

This example of "data mining" is the future, if not the present, of law enforcement.[1] Data mining both offers enormous possibilities for law enforcement and national security — in the scenario above, the fertilizer and truck could have been intended for a far less innocuous use — and radically undermines the notion that one's interests, affiliations, and legal activities can be kept private from the government. Such concerns have led to significant public debate over the proper scope of surveillance, prompted in particular by Edward Snowden's recent disclosures.

Yet despite the obvious privacy implications of data mining, traditional Fourth Amendment doctrine offers relatively little to help constrain such activity. The Supreme Court has held that one cannot have a reasonable expectation of privacy in information that is given to third parties[2] or made accessible to the public.[3] In the modern era, this doctrine covers an enormous amount of activity: commercial interactions are known to credit card companies;[4] financial records are in the hands of banks;[5] phone calls and emails entail offering telecommunications companies the numbers and addresses necessary to route the information properly;[6] and even a cell phone's location may

---

[1] *See* 1 WAYNE R. LAFAVE, SEARCH & SEIZURE § 2.7(e) (5th ed. 2012).

[2] Smith v. Maryland, 442 U.S. 735, 743–45 (1979).

[3] California v. Greenwood, 486 U.S. 35, 40–41 (1988).

[4] *See* United States v. Phibbs, 999 F.2d 1053, 1077 (6th Cir. 1993) ("It is evident, however, that Rojas did not have both an actual and a justifiable privacy interest in any of these materials, including his credit card statements and telephone records.").

[5] *See* United States v. Miller, 425 U.S. 435, 440, 442–43 (1976).

[6] *See* United States v. Forrester, 512 F.3d 500, 510 (9th Cir. 2008) (considering "the constitutionality of computer surveillance techniques that reveal the to/from addresses of e-mail messages, the IP addresses of websites visited and the total amount of data transmitted to or from an account," and concluding that "the surveillance techniques the government employed here are con-

be known to the phone company at all times.[7] Such information, under extant Supreme Court doctrine, arguably falls outside the scope of the Fourth Amendment's protections. Accordingly, the government can compile and analyze it largely free from constitutional scrutiny.[8]

Commentators have responded to this apparent deficiency by suggesting that individual sources of information might be protected, Facebook and other social networks being one widely debated example.[9] But the reality of modern data mining is that removing isolated sources from the flood of "public" information will do little to stop the government from divining a detailed portrait from the information that remains available. Whatever restraints are imposed on the collection of individual sources, it continues to be true that "most government data mining today occurs in a legal vacuum outside the scope of the Fourth Amendment and without a statutory or regulatory framework."[10]

Taking another tack, five Justices of the Supreme Court have signaled a willingness to move away from the piece-by-piece analysis toward a "mosaic theory" of the Fourth Amendment.[11] In *United States v. Jones*,[12] the majority decided that long-term surveillance via a GPS beacon attached to a car bumper constituted a search due to the physical trespass upon the bumper.[13] Yet Justice Sotomayor concurring and Justice Alito — joined by Justices Ginsburg, Breyer, and Kagan — concurring in the judgment suggested that the collection of sufficiently large amounts of information might amount to a search (thus implicating the Fourth Amendment) regardless of physical trespass.[14] Howev-

---

stitutionally indistinguishable from the use of a pen register that the Court approved in *Smith* [*v. Maryland*]").

[7] *See In re* Application of the U.S. for Historical Cell Site Data, 724 F.3d 600, 615 (5th Cir. 2013) (allowing the collection of cell site data under the third-party doctrine).

[8] *See* ACLU v. Clapper, 959 F. Supp. 2d 724, 752 (S.D.N.Y. 2013) (permitting mass NSA collection of telephonic metadata). *But see* Klayman v. Obama, 957 F. Supp. 2d 1, 36–37 (D.D.C. 2013) (finding NSA data mining would likely constitute a Fourth Amendment violation).

[9] *See, e.g.*, Monu Bedi, *Facebook and Interpersonal Privacy: Why the Third Party Doctrine Should Not Apply*, 54 B.C. L. REV. 1 (2013).

[10] Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435, 437 (2008); *see id.* at 436–37. There are statutory constraints on both public-sector and private-sector collection and use of data, consisting primarily of the Fair Credit Reporting Act of 1970, 15 U.S.C. § 1681–1681x (2012), and the Privacy Act of 1974, 5 U.S.C. § 552a (2012). *See* Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection*, 2006 U. ILL. L. REV. 357, 359–62. But these laws were passed over four decades ago; perhaps unsurprisingly, "this framework is riddled with exceptions and shunted with limitations." *Id.* at 359.

[11] *See* Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 313 (2012) (discussing United States v. Jones, 132 S. Ct. 945 (2012), in which five Justices signed concurring opinions supporting the D.C. Circuit's application of this theory). At least one court has used *Jones* to rule against mass NSA surveillance. *See Klayman*, 957 F. Supp. 2d at 36.

[12] 132 S. Ct. 945.

[13] *Id.* at 950–53.

[14] *See id.* at 956–57 (Sotomayor, J., concurring); *id.* at 964 (Alito, J., concurring in the judgment).

er, the question as framed in *Jones* — to what degree a given type of information can be analyzed — is likely to offer little guidance to courts struggling to determine the validity of investigations using multiple types of rapidly evolving information-gathering techniques. More promisingly, scholars have noted the possibility of building legal protections into algorithms and databases in order to protect privacy while still enhancing law enforcement capabilities.

This Note argues that a properly designed algorithmic search, with features corresponding to the Fourth Amendment's dog-sniffing doctrine, can offer a potential constitutional solution to the privacy pitfalls of data mining. In a series of cases the Court has stated that the use of drug-sniffing dogs does not constitute a search under certain conditions. From these cases the following requirements can be identified: that the dogs access the scents without intruding into a constitutionally protected area, that they recognize only illegal activity, that humans do not see any private information until probable cause has been established by the dog's bark, and that the dogs have a low false-positive rate. These features map roughly onto the characteristics of a well-designed algorithmic search.

This Note begins by discussing data mining: its definition, its utility, and the threat it presents to traditional notions of privacy. Part II analyzes the extent to which data mining can be regulated under established Fourth Amendment doctrine, agreeing with the scholarly consensus that it largely falls outside the traditional scope of a search. Part III explores some of the alternatives that have been put forward, finding some promise in regulating access to certain types of information and less in the mosaic theory hinted at in recent cases. Part IV presents this Note's alternative, arguing that it flows logically from dog-sniff doctrine and answers the most serious objections to data mining. Part V concludes.

I. DATA MINING'S PROMISES AND PITFALLS

The quantity of information collected about U.S. citizens, both privately and publicly, is expanding at a prodigious rate.[15] The government has direct access to an enormous amount of information collected by various agencies: payroll records, political contribution disclosure regimes, birth certificates, marriage licenses, and more.[16] The federal government maintained more than 2,000 databases over a

[15] *See* LAFAVE, *supra* note 1, § 2.7(e).

[16] *See* Newton N. Minow & Fred H. Cate, *Government Data Mining*, *in* THE MCGRAW-HILL HOMELAND SECURITY HANDBOOK 1133, 1134 (David G. Kamien ed., 2d ed. 2012).

decade ago,[17] a number that surely understates today's figures.[18] In addition to the numerous public-sector sources of information, the private sector has amassed considerable information about consumers.[19] Part of this trend can be traced to the increasing number of interactions and transactions that occur online and electronically, as email replaces mail, Amazon.com replaces storefronts, credit cards replace cash, and Facebook replaces conversation. This proliferation of available data — combined with the demand for such data from both public and private sources[20] — has led to "the creation of a new industry: the database industry."[21] This industry "provides data to companies for marketing, to the government for law enforcement purposes, to private investigators for investigating individuals, to creditors for credit checks, and to employers for background checks."[22]

And what is being done with these data? They are examined, either by people or by algorithms, for patterns of useful information in a process termed "data mining." Data mining, for the purposes of constitutional analysis of government surveillance, can be defined as "searches of one or more electronic databases of information concerning U.S. persons, by or on behalf of an agency or employee of the government."[23] Data mining, of course, can also be carried out by private parties — one famous example involved Target's analyzing the shopping habits of its customers to identify those who had recently become pregnant, and preemptively targeting them with baby-product advertisements.[24] More sophisticated uses involve massive databases compiled by both governments and private companies from a wide variety of sources that can be used to target advertising or law enforcement resources.[25]

---

[17] Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1403 (2001).

[18] One indication is that federal government spending on gathering and maintaining statistical databases, excluding the census, has increased from $3.79 billion in 2001, *see* OFFICE OF MGMT. & BUDGET, STATISTICAL PROGRAMS OF THE UNITED STATES GOVERNMENT: FY 2003, at 7, http://www.whitehouse.gov/sites/default/files/omb/inforeg/03statprog.pdf [http://perma.cc/H8XW-VGRX], to $6.29 billion in 2012, *see* OFFICE OF MGMT. & BUDGET, STATISTICAL PROGRAMS OF THE UNITED STATES GOVERNMENT: FY 2014, at 10, http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2014.pdf [http://perma.cc/T36-EP4S], an increase of 28% adjusted for inflation.

[19] *See generally* ROBERT O'HARROW, JR., NO PLACE TO HIDE (2005).

[20] *See* Lee Tien, *Privacy, Technology and Data Mining*, 30 OHIO N.U. L. REV. 389, 389–90 (2004).

[21] Solove, *supra* note 17, at 1407.

[22] Solove & Hoofnagle, *supra* note 10, at 363.

[23] TECH. & PRIVACY ADVISORY COMM., DEP'T OF DEFENSE, SAFEGUARDING PRIVACY IN THE FIGHT AGAINST TERRORISM xiii (2004), http://epic.org/privacy/profiling/tia/tapac_report.pdf [http://perma.cc/ZEJ6-GN45].

[24] *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES, Feb. 16, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[25] *See* LAFAVE, *supra* note 1, § 2.7(e).

More specifically, data mining can be grouped into two broad categories: "subject-based," which involves pulling together and analyzing information about a previously identified individual, and "pattern-based," which involves analyzing information on nonsuspect individuals to identify patterns of transactions or behaviors that correlate with suspect activity.[26] While subject-based data mining may raise constitutional concerns of its own, this Note focuses primarily on pattern-based data mining. Such data mining in the absence of individualized suspicion differs in kind, not merely degree, from traditional government investigatory techniques.

Data mining holds undeniable promise for law enforcement: it can "turn[] low-level data, usually too voluminous to understand, into higher forms (information or knowledge) that might be more compact (for example, a summary), more abstract (for example, a descriptive model), or more useful (for example, a predictive model)."[27] Just as Target was able to predict pregnancy, a government could in theory identify transactions indicative of tax fraud or drug dealing, or of terrorist attacks in the making.[28] Data mining's greatest advantage over traditional forms of surveillance is that it does not require ex ante individualized suspicion: law enforcement could identify a past (or even future) wrongdoer whom the government would otherwise never have suspected. In theory, law enforcement could also become more efficient, in terms of both cost and burden on citizens. If police could identify criminals through data mining and disrupt embryonic terror attacks, one could envision a future where passengers can wear shoes through airport security. While there are debates about exactly how effective data mining can be for law enforcement and national security purposes,[29] law enforcement and national security agencies are rapidly

---

[26] *See* Minow & Cate, *supra* note 16, at 1134–35.

[27] K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 1, 22 (2003).

[28] *See id.* at 3 n.1 (citing Charles Piller & Eric Lichtblau, *FBI Plans to Fight Terror with High-Tech Arsenal*, L.A. TIMES, July 29, 2002, at A1 ("By Sept. 11, 2011, the FBI hopes to use artificial-intelligence software to predict acts of terrorism the way the telepathic 'precogs' in the movie 'Minority Report' foresee murders before they take place.")). That is not to say that law enforcement will actually achieve high-percentage predictive capabilities in the near future, or ever.

[29] *See* JEFF JONAS & JIM HARPER, EFFECTIVE COUNTERTERRORISM AND THE LIMITED ROLE OF PREDICTIVE DATA MINING, (Cato Inst. Policy Analysis No. 584, 2006), http://object.cato.org/sites/cato.org/files/pubs/pdf/pa584.pdf [http://perma.cc/DTY9-VHNB]. *Compare* Holman W. Jenkins, Jr., *Can Data Mining Stop the Killing?*, WALL ST. J. (July 24, 2012, 6:53 PM), http://online.wsj.com/article/SB10000872396390443570904577546671693245302.html [http://perma.cc/G7JL-KY7N] (asking whether the Pentagon's now-defunct Total Information Awareness program could have stopped the Aurora theater shooting), *with* Shane Harris, *Data Mining Would Not Have Stopped the Aurora Shooting*, WASHINGTONIAN: CAPITAL COMMENT (July 26, 2012), http://www.washingtonian.com/blogs/capitalcomment/media/data-mining-would-not-have-stopped-the-aurora-shooting.php [http://perma.cc/9C79-3NBC] (answering "no").

expanding their efforts and capabilities to gather information and analyze it on a mass scale.[30]

Yet data mining's promise for law enforcement comes paired with significant privacy concerns. The privacy concerns attendant to data mining (as opposed to information-gathering more generally) can be grouped into "those that arise from the aggregation (or integration) of data and those that arise from the automated analysis of data that may not be based on any individualized suspicion."[31] The first concern is that discussed by Justice Sotomayor in *Jones* — "GPS monitoring generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations."[32] This concern becomes even more severe when GPS data is combined with credit card transactions, online activities, and other forms of data.

The second concern, meanwhile, is rooted in the unnerving fact that this intimate, invasive surveillance is targeted at *everyone*. More traditional government surveillance may not require a warrant until reaching the level of a search; yet presumably law enforcement is not investing the resources necessary to surveil round the clock and track down every piece of information without at least some whiff of wrongdoing. The average citizen can take comfort in the assumption that she will not incur such close scrutiny, a comfort that is not afforded by pattern-based data mining.

These concerns have most recently come to the fore due to the revelations of former National Security Agency (NSA) contractor Edward Snowden. Snowden's revelations about the scope of NSA surveillance have prompted a wave of privacy concerns and a renewed debate around the tradeoffs between privacy and security attendant in data mining.[33] It is unclear as of yet what, if any, legislative action will result from these revelations; in the meantime, courts have struggled with the constitutional implications of such programs.

II. DATA MINING UNDER THE FOURTH AMENDMENT

Fourth Amendment doctrine rests upon two assumptions that data mining exposes as particularly ill-suited to the modern age: that physical intrusions will correspond to the most serious invasions of privacy, and that the inability of government to invade privacy on a mass scale

---

[30] *See* Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317, 317–20 (2008) (describing government efforts across multiple departments).

[31] Taipale, *supra* note 27, at 57.

[32] United States v. Jones, 132 S. Ct. 945, 955 (2012) (Sotomayor, J., concurring).

[33] *See generally* Stephen I. Vladeck, *Big Data Before and After Snowden*, 7 J. NAT'L SECURITY L. & POL'Y 333 (2014).

will offer practical obscurity. The first assumption takes root in the fact that opening one's mail, entering one's property, or rooting through one's belongings all involve clear lines of physical intrusion that courts can easily police. Yet as technology advances, the policing of physical intrusion starts to look very much like the Maginot line: impregnable against frontal assault while far more serious invasions of privacy flow around it unimpeded. The Court is thus adamant that to set foot on private property is to trigger a search,[34] yet relatively unconcerned about helicopters (and soon drones, no doubt) hovering close above one's property with high-resolution cameras.[35] Such vigilance offers little comfort in a world where one's intimate transactions occur in a space where no physical intrusion is required to access them, and the Court has begun to react with trepidation to the conflict between these traditional assumptions and the modern world.

A. Traditional Fourth Amendment Doctrine

Two major cases in the Fourth Amendment canon have left a vast amount of data constitutionally unprotected. First, the Supreme Court declared in *California v. Greenwood*[36] that one does not have a privacy interest in garbage placed out on the street for collection,[37] and more generally that the Fourth Amendment does not protect that which "could have been observed by any member of the public."[38] Thus one's public movements and actions, prior to *Jones*, were thought not to receive Fourth Amendment protection.[39]

Second, and more problematic to scholars,[40] the Court stated in *Smith v. Maryland*[41] that an individual has no "legitimate expectation of privacy in information he voluntarily turns over to third parties."[42] The paradigmatic examples of this principle are bank records[43] and

---

[34] *See* Florida v. Jardines, 133 S. Ct. 1409, 1417 (2013).

[35] *See* California v. Ciraolo, 476 U.S. 207, 215 (1986). One could also note, as the concurrences did in *Jones*, that Justice Scalia's stalwart defense of car owners' right to have their bumpers free of small GPS devices is all but useless if the government can accomplish the same ends by simply requesting GPS data from the companies that provide mapping services. *See Jones*, 132 S. Ct. at 955 (Sotomayor, J., concurring) ("With increasing regularity, the Government will be capable of duplicating the monitoring undertaken in this case by enlisting factory- or owner-installed vehicle tracking devices or GPS-enabled smartphones."); *id.* at 961 (Alito, J., concurring in the judgment).

[36] 486 U.S. 35 (1988).

[37] *Id.* at 40.

[38] *Id.* at 41.

[39] *See* United States v. Knotts, 460 U.S. 276, 281 (1983).

[40] *See generally, e.g.*, Stephen E. Henderson, *Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too*, 34 PEPP. L. REV. 975 (2007).

[41] 442 U.S. 735 (1979).

[42] *Id.* at 743–44.

[43] United States v. Miller, 425 U.S. 435, 436–40 (1976).

telephone numbers dialed.[44] Today, this third-party doctrine appears to extend as far as recording one's "IP address, to/from address for e-mails, and volume sent from the account."[45] While the contents of emails might receive protection,[46] the lines of Fourth Amendment searches as set by the Court's application of analog doctrines bear virtually no resemblance to society's current expectations of privacy.[47]

The cumulative effect of the public exposure and third-party doctrines renders data mining largely "outside the scope of the Fourth Amendment."[48] While there are statutory restrictions on certain types of surveillance, most notably the Stored Communications Act,[49] the Fourth Amendment leaves unprotected any information that has fallen or could legally fall into the hands of a private third party. Accordingly, a staggering amount of information generally considered quite personal can be collected with limited constitutional restriction.

These two doctrines interact problematically with another core assumption of the Fourth Amendment: that law enforcement has limited resources and cannot be in all places at all times. This assumption has meant that the Court has yet to recognize that both the extent to which data are analyzed and the scope of their collection have constitutional implications. Courts have traditionally assumed a degree of practical obscurity: even if one cannot guarantee the privacy of one's transactions against the watchful eye of the state, one can reasonably expect that government agents will not follow one's public movements, collect receipts at every vendor one visits, and check the address on every letter one sends or receives.[50]

Courts have most fully articulated this principle in the context of public movement. Discussing information-gathering police stops, the Court has relied upon "limited police resources" along with other practical constraints to inhibit "an unreasonable proliferation of police checkpoints."[51] Judge Posner, meanwhile, has distinguished between the police's ability to follow a single driver through public streets and

---

[44] *Smith*, 442 U.S. at 742.

[45] Orin S. Kerr, *Applying the Fourth Amendment to the Internet: A General Approach*, 62 STAN. L. REV. 1005, 1027 (2010); *see also id.* at 1027–28 (discussing United States v. Forrester, 512 F.3d 500 (9th Cir. 2008)).

[46] Stephen E. Henderson, *The Timely Demise of the Fourth Amendment Third Party Doctrine*, 96 IOWA L. REV. BULL. 39, 40–42 (2011) (discussing recent lower court decisions and the Supreme Court's rather inconclusive decision in City of Ontario v. Quon, 560 U.S. 746 (2010)).

[47] *See* Slobogin, *supra* note 30, at 333–36 (comparing survey results about attitudes toward different types of searches with the extent of current Fourth Amendment protections).

[48] Cate, *supra* note 10, at 437.

[49] 18 U.S.C. §§ 2701–2712 (2012).

[50] *See* Taipale, *supra* note 27, at 58–60.

[51] Illinois v. Lidster, 540 U.S. 419, 426 (2004).

the possibility of mass observation through technology (and even analysis of movement patterns via algorithm).[52]

While the Court has recognized that the elimination of such practical barriers to information gathering can independently raise Fourth Amendment concerns,[53] it continues to rely upon the default assumption of practical obscurity. In fact, such an assumption underlies the fundamental mode of Fourth Amendment analysis: each step in a search is to be analyzed independently for any constitutional violation, regardless of the number of steps or searches put together.[54] While individualized analysis might make sense when each element of a search requires an investment of significant resources, it seems hopelessly outdated when thousands of micro-searches can be effortlessly amalgamated.[55]

B. Recent Conflicts Between Technology and Fourth Amendment Doctrine

The Court is hardly unaware of the challenges that technological development has posed to its traditional Fourth Amendment assumptions. Though the Court has yet to encounter data mining directly,[56] in a series of recent cases it has expressed trepidation about uninhibited adoption of technologically dated Fourth Amendment precedents.

First, the Court has hesitated to allow search of email stored on a third party's servers. In *City of Ontario v. Quon*,[57] the Court was faced with the question of whether an employee could have a reasonable expectation of privacy in text messages stored on a government employer's servers.[58] Yet rather than address the question head on,

---

[52] *See* United States v. Garcia, 474 F.3d 994, 997–98 (7th Cir. 2007) ("One can imagine the police affixing GPS tracking devices to thousands of cars at random, recovering the devices, and using digital search techniques to identify suspicious driving patterns. . . . It would be premature to rule . . . that it could not be a search because it would merely be an efficient alternative to hiring another 10 million police officers to tail every vehicle on the nation's roads.").

[53] *See* U.S. Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 764 (1989) (recognizing that "the compilation of otherwise hard-to-obtain information alters the privacy interest implicated by disclosure of that information").

[54] Kerr, *supra* note 11, at 315–16 (describing the traditional step-by-step approach).

[55] *See* Joseph S. Fulda, *Data Mining and Privacy*, 11 ALB. L.J. SCI. & TECH. 105, 106 (2000) (noting the erosion of cost-benefit restraints on law enforcement due to technological advances). *But see* Orin Kerr, *Debate: Metadata and the Fourth Amendment — A Reply to Jennifer Granick*, JUST SECURITY (Sept. 23, 2013, 3:30 PM), http://justsecurity.org/1009/debate-metadata-fourth-amendment-reply-jennifer-granick [http://perma.cc/864-4G6S] (arguing that the scope of a surveillance program is constitutionally irrelevant, because "the Fourth Amendment is about *individual* rights, protecting each person from unreasonable searches and seizures").

[56] The nearest the Court has come to considering mass warrantless surveillance was in *Clapper v. Amnesty International USA*, 133 S. Ct. 1138 (2013), which the Court dismissed on standing grounds. *See id.* at 1143.

[57] 130 S. Ct. 2619 (2010).

[58] *Id.* at 2624.

the Court ruled that the search was reasonable regardless of the employee's privacy interest.[59] In explaining the Court's reticence, Justice Kennedy explained that "[t]he Court must proceed with care . . . . The judiciary risks error by elaborating too fully on the Fourth Amendment implications of emerging technology before its role in society has become clear."[60] While this case could be read as a simple application of the canon of constitutional avoidance, the Court has often cast such modesty aside in the field of criminal procedure.[61] Rather, *City of Ontario* may indicate that the Court is reluctant to follow *Smith* all the way down the rabbit hole when it comes to electronic communications.

Second, in *United States v. Jones*, the Court confronted the use of a GPS tracking device to surveil a suspect for four weeks. Decades earlier, in *United States v. Knotts*,[62] the Court had held that the use of a locating "beeper" was constitutionally permissible because "[a] person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another."[63] The *Jones* majority distinguished this seemingly controlling precedent by finding that while in *Knotts* the suspect had voluntarily placed a bugged package in his car, in *Jones* the government trespassed upon the suspect's rear bumper in placing the device.[64] Yet for the five Justices concurring, it was not the origin of the device but the extent of its information gathering that was most troubling. Justice Alito dismissed the attachment of the device as "trivial," and argued that the length of the surveillance passed an as-yet unidentified threshold marking the bounds between a search and a non-search.[65] Justice Sotomayor went even further, specifically calling into question the viability of the third-party doctrine in

> the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks. People disclose the phone numbers that they dial or text to their cellular providers; the URLs that they visit and the e-mail addresses with which they correspond to their Internet service providers; and the books, groceries, and medications they purchase to online retailers.[66]

Last Term, the Court struck an even more direct blow at technological neutrality — the notion that the Fourth Amendment should trans-

---

[59] *See id.* at 2632.

[60] *Id.* at 2629.

[61] *See The Supreme Court, 2011 Term — Leading Cases*, 126 HARV. L. REV. 176, 241 (2012) ("[T]his restraint is strikingly absent in cases presenting questions of constitutional criminal procedure.").

[62] 460 U.S. 276 (1983).

[63] *Id.* at 281.

[64] United States v. Jones, 132 S. Ct. 945, 951–52 (2012).

[65] *Id.* at 961–64 (Alito, J., concurring in the judgment).

[66] *Id.* at 957 (Sotomayor, J., concurring).

late seamlessly from the analog to the digital.[67] In *Riley v. California*,[68] the Court unanimously refused to extend the traditional search-incident-to-arrest exception — by which arresting officers could rifle through the effects of an arrestee without Fourth Amendment scrutiny[69] — to the search of an arrestee's cell phone. Chief Justice Roberts explained that to compare the search of a cell phone to that of a wallet or a purse "is like saying a ride on horseback is materially indistinguishable from a flight to the moon. . . . Modern cell phones, as a category, implicate privacy concerns far beyond those implicated by the search of a cigarette pack, a wallet, or a purse. . . . [A]ny extension of that reasoning to digital data has to rest on its own bottom."[70]

These cases suggest that the Court is aware that modern surveillance technologies represent a problem for traditional Fourth Amendment doctrine, but is still casting about for a solution that might prove workable in the context of data mining. In the next Part, this Note examines the alternatives that have been put forward.

III. FOURTH AMENDMENT ALTERNATIVES

Several proposals have been floated to address the mounting unease with the mass collection and analysis of data that, while (arguably) innocuous in pieces, in combination can reveal a discomfiting amount about a person's life. Broadly speaking, these proposals can be grouped into three categories: those that restrict what types of information can be gathered, those that restrict how much of it can be put together, and those that restrict how it can be analyzed. Though the first and second categories are important, this Note focuses on the third category as offering the most potential for systematic judicial regulation of data mining.

A. What Can Be Collected?

Much of the scholarly attention has focused on restricting the types of data that can be collected. Some critics have attacked the third-party doctrine directly, arguing either that the entire edifice is built upon a mistake,[71] or that it should distinguish information that is exposed to a third party only by passing through an automated conduit to another private party (so that, for example, emails that pass through

---

[67] Professor Orin Kerr identifies technological neutrality as a core value of Fourth Amendment doctrine, and the third-party doctrine as essential to preserving this neutrality. *See* Orin S. Kerr, *The Case for the Third-Party Doctrine*, 107 MICH. L. REV. 561, 579–81 (2009).

[68] 134 S. Ct. 2473 (2014).

[69] *See* United States v. Robinson, 414 U.S. 218, 224 (1973).

[70] *Riley*, 134 S. Ct. at 2488–89.

[71] *See* Henderson, *supra* note 46, at 40.

third-party servers would not lose their protected character).[72] Without wading too deep into the continuing vitality of the third-party doctrine, it is worth noting that such proposals run squarely into *Smith*, which solidified the third-party doctrine as applicable to information collected in the course of automated communications.[73] Other scholars have focused on those types of information that implicate other constitutional interests, such as associational or interpersonal privacy. The question of how much privacy one is entitled to in the information one posts on Facebook has generated its own small field of constitutional scholarship.[74] Additional scholarship has focused on location tracking, arguing that the pervasive surveillance of one's public movements could offend the Constitution.[75]

Regulation of what can legitimately be collected is undoubtedly important. Even if legal restrictions are placed on the scope of data analysis, it would offend the Constitution if the inputs into a data-mining program included intimate conversations within the marital bedroom. Yet analyzing each source of information smacks of attempting to hold back the flood by plugging each leak in the dam as it appears. Justice Sotomayor noted in *Jones* the enormous amount of personal information that could be garnered from GPS tracking alone.[76] Yet at the same time, excluding location data would hardly prevent the government from generating much the same record by looking solely at one's email exchanges, browser history, or credit card transactions: "[I]t will often be unnecessary for the government to track us, because for most of us much of our lives are already described in transactional databases."[77] Unless every meaningful source of information is to be regulated, a more systematic approach is needed.

B. How Much Can Be Put Together?

An alternative (or additional) approach to the regulation of data mining is to look not merely at sources, but at the amount of information that is accumulated. Professor Orin Kerr describes this as the "mosaic theory," and notes its endorsement by the D.C. Circuit in

---

[72] *See* Matthew Tokson, *Automation and the Fourth Amendment*, 96 IOWA L. REV. 581, 609–11 (2011).

[73] Smith v. Maryland, 442 U.S. 735, 744–45 (1979) ("We are not inclined to hold that a different constitutional result is required because the telephone company has decided to automate.").

[74] *See, e.g.*, Bedi, *supra* note 9.

[75] *See, e.g.*, Reeve Wood, Comment, *The Prolonged Arm of the Law: Fourth Amendment Principles, the* Maynard *Decision, and the Need for a New Warrant for Electronic Tracking*, 64 ME. L. REV. 285 (2011).

[76] United States v. Jones, 132 S. Ct. 945, 955–56 (2012) (Sotomayor, J., concurring).

[77] Tien, *supra* note 20, at 400.

United States v. Maynard[78] and by the concurrences in *Jones*. The mosaic theory "considers whether a set of nonsearches aggregated together amount to a search because their collection and subsequent analysis creates a revealing mosaic."[79] Such an approach accords with our intuitions and expectations about privacy: the government may be entitled to examine a particular commercial transaction, or to find out where a suspect is at a given moment, but should not be able to piece together her entire life without first seeking a warrant.

Yet Kerr is right to note the significant difficulties involved in setting forth a predictable standard for the mosaic theory.[80] First, the three relevant opinions — Judge Ginsburg's in *Maynard*, Justice Sotomayor's *Jones* concurrence, and Justice Alito's *Jones* concurrence — put forward three divergent variations of the mosaic theory test, each different in important respects from the others.[81] More troublingly, it is difficult to see how any standard could reliably apply either within types of surveillance (three days of GPS tracking is acceptable, but is four days too many?) or across types (bank records are okay, as are email addresses, but do the two combined create a search?). Unless one imagines each type of nonsearch being assigned a point value that can accumulate to a search, the mosaic theory is not likely to lend itself to stable solutions, but rather to frustrate equally both government investigators and privacy advocates.

C. What Can Be Done with the Data?

If regulation of individual sources of information at the point of collection is insufficient (though indispensable), and regulation of the gross extent of analysis is likely to result in endless confusion, perhaps one should examine the method of analysis. Some scholars who focus on the method of analysis have looked to the distinction identified above between subject-based data mining (the examination of accumulated information on a pre-identified individual) and pattern-based data mining (the suspicionless examination of large numbers of individuals for indicative patterns of behavior). While subject-based data mining may be a logical extension of ordinary investigative techniques, pattern-based data mining has drawn particular ire: such analysis, divorced from particularized suspicion, is viewed as hostile to both "the constitutional presumption of innocence and the Fourth Amendment principle that the government must have individualized suspicion be-

---

[78] 615 F.3d 544 (D.C. Cir. 2010), *aff'd in part sub nom.* United States v. Jones, 132 S. Ct. 945 (2012).

[79] Kerr, *supra* note 11, at 320.

[80] *Id.* at 330–33; *see also* Benjamin M. Ostrander, Note, *The "Mosaic Theory" and Fourth Amendment Law*, 86 NOTRE DAME L. REV. 1733, 1766 (2011).

[81] *See* Kerr, *supra* note 11, at 330.

fore it can conduct a search."[82] Yet it is precisely the ability to investigate in the absence of preexisting suspicion that offers data mining's greatest promise: the possibility of putting together disconnected facts to point the finger at a suspect whom the government would not otherwise have suspected.

Rather than throw the baby out with the bathwater, a more promising avenue is to regulate the analysis of the data in a manner that comports with constitutional principles. K.A. Taipale has discussed this possibility at length, arguing that "security with privacy can be achieved by employing value-sensitive technology development strategies that take privacy concerns into account during development, in particular, by building in rule-based processing, selective revelation, and strong credential and audit features."[83] Taipale analyzes each of these features in depth: rule-based processing allows data to be labeled and categorized in order to ensure that it will not be accessed improperly;[84] selective revelation "uses an iterative, layered structure that reveals personal data partially and incrementally in order to maintain subject anonymity";[85] and strong credentialing and audit features avoid insider abuse of information by restricting and monitoring access.[86] Taipale demonstrates the viability of these features as a technological matter and argues convincingly that they can allow data mining to accord with privacy intuitions.

Whereas Taipale focuses his attention on the feasibility and social desirability of certain features of data mining, however, this Note is more concerned with "how these particular technologies fit within the current legal structure."[87] As the *Riley* Court made clear, the adoption of privacy-enhancing protocols is to be encouraged, but constitutional scrutiny remains indispensable: "[T]he Government proposes that law enforcement agencies 'develop protocols to address' concerns raised by cloud computing. Probably a good idea, but the Founders did not fight a revolution to gain the right to government agency protocols."[88] Indeed they did not. But, as argued in Parts II and III, neither the status quo nor the solutions so far offered are likely to provide a coherent and satisfactory accommodation between competing constitutional concerns. In the next Part, this Note thus seeks to fill a gap by identifying

---

[82] James X. Dempsey & Lara M. Flint, *Commercial Data and National Security*, 72 GEO. WASH. L. REV. 1459, 1466–67 (2004).

[83] Taipale, *supra* note 27, at 1.

[84] *See id.* at 75–78.

[85] *Id.* at 79; *see id.* at 79–80.

[86] *See id.* at 80.

[87] *Id.* at 50–51.

[88] Riley v. California, 134 S. Ct. 2473, 2491 (2014) (citation omitted) (quoting Reply Brief for the United States at 14, *Riley*, 134 S. Ct. 2473 (No. 13-212), 2014 WL 1616437, at *14).

one model of constitutional oversight of data mining and demonstrating its congruence with existing Fourth Amendment doctrine.

IV. THE CRIME-SNIFFING ALGORITHM

This Part examines the Court's treatment of the use of drug- and explosive-sniffing dogs under the Fourth Amendment. While such cases have generally been relegated to a niche, the elements of the doctrine map surprisingly well onto the constitutional issues posed by data mining. Analogizing from the cases determining whether dog-sniffing creates Fourth Amendment concerns, this Note lays out the elements that a data-mining algorithm would have to satisfy: the initial search must be performed by a computer upon a database of traditionally unprotected information; the algorithm must not identify protected (noncriminal) activity; human interaction with the data must occur only after the algorithm has demonstrated probable cause; and the algorithm must have a sufficiently low false-positive rate.

A. The Constitutionality of the Drug-Sniffing Dog

In four cases spread decades apart, the Supreme Court confirmed that the use of a drug-sniffing dog, in a manner that did not involve additional intrusion beyond that already constitutionally permissible, did not constitute a search under the Fourth Amendment and that the dog's reaction could provide probable cause for a search. In *United States v. Place*,[89] the Court established this rule in upholding a sniff test of luggage pursuant to a valid *Terry* stop,[90] and in *Illinois v. Caballes*[91] the Court confirmed it with regard to a sniff test of a vehicle's exterior, again pursuant to a valid *Terry* stop.[92] Most recently, in *Florida v. Jardines*,[93] the Court held that stepping onto the curtilage of a home with a drug-sniffing dog constituted a Fourth Amendment violation.[94] The same Term, in *Florida v. Harris*,[95] the Court confirmed that "[a] sniff is up to snuff" in establishing probable cause.[96] From these cases four important features of the doctrine can be drawn: the sniff must only analyze information that is legally obtained; the sniff must only detect illegal activity; humans must not participate in any

---

[89] 462 U.S. 696 (1983).

[90] *Id.* at 706–07. In *Terry v. Ohio*, 392 U.S. 1 (1968), the Court permitted officers to briefly detain suspicious persons upon less than probable cause in order to pursue an investigation. *See* Adams v. Williams, 407 U.S. 143, 145–46 (1972) (discussing *Terry*).

[91] 543 U.S. 405 (2005).

[92] *Id.* at 408–09.

[93] 133 S. Ct. 1409 (2013).

[94] *Id.* at 1417–18.

[95] 133 S. Ct. 1050 (2013).

[96] *Id.* at 1058.

search until probable cause has been established by the sniff; and the sniff must have a low false-positive rate.

The first crucial feature is that the dog, because it does not physically intrude into the bag or car, conducts its detection from a nontrespassory vantage point. In *Place*, the Court identified this feature by pointing out that a "'canine sniff' by a well-trained narcotics detection dog . . . does not require opening the luggage."[97] The *Caballes* Court agreed, applying the logic to a sniff of a car's exterior.[98] In *Jardines*, the fact that a government agent had stepped onto the property with the drug-sniffing dog provided the critical distinction from *Place* and *Caballes*, as the Court deemed the activity a search on trespass grounds.[99] Put another way, dog sniffs are permissible so long as they gather data where it has emerged from a constitutionally protected space into a constitutionally *un*protected space. That the information is no longer technically within the home is not itself sufficient — in *Kyllo v. United States*[100] the Court found a search where police used a thermal imaging device from across the street that detected heat radiating from the home[101] — but it is necessary that obtaining the information to be analyzed not involve an independent constitutional violation.

The second important feature of dog sniffing identified by the Court is that dogs are trained only to react to illegal activity. As the Court stated in *Caballes*, "governmental conduct that *only* reveals the possession of contraband 'compromises no legitimate privacy interest.'"[102] The Court relied on a previous case holding that chemical analysis of white powder for the presence of cocaine did not constitute a search.[103] This principle seems simple enough, but there is a critical distinction to be drawn out between what the dog detects and what the dog reacts to. The dog (and even more so the chemical field test) only reacts in a binary manner: drugs or no drugs (or perhaps drugs and/or explosives, or neither). However, the dog and the test necessarily encounter scents and substances that are not only innocent, but po-

---

97 United States v. Place, 462 U.S. 696, 707 (1983).

98 Illinois v. Caballes, 543 U.S. 405, 409 (2005).

99 *Jardines*, 133 S. Ct. at 1417. The Court notably passed up the opportunity to clarify whether a dog sniff from *outside* the property — absent a trespass — constituted a search. *Compare* United States v. Thomas, 757 F.2d 1359, 1367 (2d Cir. 1985) (finding a search), *with* United States v. Colyer, 878 F.2d 469, 477 (D.C. Cir. 1989) (opposite). Justice Alito, joined by Chief Justice Roberts and Justices Kennedy and Breyer, would not have found a search. *See Jardines*, 133 S. Ct. at 1424–26 (Alito, J., dissenting). Justice Kagan, joined by Justices Ginsburg and Sotomayor, would have found a search. *See id.* at 1418–20 (Kagan, J., concurring).

100 533 U.S. 27 (2001).

101 *Id.* at 31–40.

102 543 U.S. at 408 (quoting United States v. Jacobsen, 466 U.S. 109, 123 (1984)).

103 *Jacobsen*, 466 U.S. at 123 ("A chemical test that merely discloses whether or not a particular substance is cocaine does not compromise any legitimate interest in privacy.").

tentially highly personal: a dog trained to do so could surely identify the scent of one's soiled undergarments or a mistress's perfume, while a field test could as easily identify one's medication. Yet because the dog and field test are only trained and designed to respond in distinctive ways to specific objects,[104] any private information they come across is meaningless to them. It is thus crucial that, while the dog may encounter private activity, it only recognizes and reports illegal activity.

The third important feature of the dog sniff is its place in the overall search process: that is, the dog must establish probable cause before a human can encounter any private information. This point was critical in *Place*, where Justice O'Connor noted that a dog sniff "does not expose noncontraband items that otherwise would remain hidden from public view, as does, for example, an officer's rummaging through the contents of the luggage. Thus, the manner in which information is obtained through this investigative technique is much less intrusive than a typical search."[105] The Court recently reaffirmed the importance of the dog's function in establishing probable cause in *Harris*, holding that "a probable-cause hearing focusing on a dog's alert should proceed much like any other."[106]

These last two features — that the dog reacts only to the presence of contraband and that a human does not become involved until probable cause is established — depend on the fourth feature of drug-sniffing dogs: a low false-positive rate. Justice Stevens was careful to note this feature of the drug-sniffing dog in *Caballes*.[107] Justice Souter vehemently contested whether the dogs actually performed as advertised, finding a range of reported false-positive rates between seven and sixty percent.[108] While one might argue that "[t]he infallible dog . . . is a creature of legal fiction,"[109] the point remains that a dog with a high false-positive rate is legally distinct from one with a low false-positive rate.[110] Thus, in *Harris*, the Court acknowledged that a dog's record of accuracy and reliability are critical to its utilization in establishing probable cause.[111]

Whatever the empirical truth of the propositions, the dog-sniffing cases suggest that a sniff properly should analyze only legally obtained

---

104 *See Jardines*, 133 S. Ct. at 1418 (Kagan, J., concurring).

105 United States v. Place, 462 U.S. 696, 707 (1983).

106 Florida v. Harris, 133 S. Ct. 1050, 1058 (2013).

107 *See* 543 U.S. at 409.

108 *See id.* at 411–12 (Souter, J., dissenting).

109 *Id.* at 411.

110 Justice Stevens suggested that "an erroneous alert, in and of itself, [does not] reveal[] any legitimate private information," *id.* at 409 (majority opinion), but this is quite plainly beside the point. The issue is what the dog's handler does *after* the erroneous alert.

111 133 S. Ct. at 1053–58.

information and detect only illicit activities; that officers should not act before probable cause has been provided by a dog's alert; and that the sniff should seldom produce false positives.

B. The Translation to Data Mining

In its dog-sniffing cases, the Court described the drug-sniffing dog as "sui generis,"[112] but as Justice Kagan noted, the highly trained dogs at issue are no different than any other "specialized device for discovering objects not in plain view."[113] The underlying logic of the dog-sniff cases fits neatly with the issues posed by pattern-based data mining. Given courts' fondness for reasoning by analogy in Fourth Amendment cases involving technological developments,[114] it should be possible to design an automated search that replicates the core features identified in the dog-sniff cases: analysis only of legally obtained information; exclusive focus on detecting illegal activity; no human observation without prior probable cause; and low error rates.

The first feature — that the database contain only legally obtained information — pertains to the database rather than the algorithm, and it is important to note that this feature falls into the category of "what can be collected" discussed above.[115] As acknowledged, the inquiry into what individual data points are or are not private is critical. One hopes that the government is already in compliance with this feature as reflected in current Fourth Amendment doctrine. A promising indicator is that, based on public accounts, the NSA program drawing the most scrutiny analyzes telephonic metadata (numbers dialed, length of call, etc.) rather than the contents of calls themselves.[116] While the dataset available to NSA algorithms vastly outstrips that available to drug-sniffing dogs, the two are comparable in legal terms. In the dog-sniff context, the information is legally obtained so long as the police and dog do not trespass while obtaining the scents. Similarly, a properly designed algorithm would analyze information that has been turned over or exposed to third parties, rather than intrude into personal computers or the content of email. Accordingly, so long as the database subject to pattern-based data mining only includes information that is gathered in accordance with current constitutional doc-

---

[112] *Caballes*, 543 U.S. at 409 (emphasis omitted) (quoting United States v. Place, 462 U.S. 696, 707 (1983)) (internal quotation marks omitted).

[113] Florida v. Jardines, 133 S. Ct. 1409, 1418 (2013) (Kagan, J., concurring).

[114] *See* Kerr, *supra* note 45, at 1027–28.

[115] *See supra* Part III.A, pp. 701–02.

[116] *See* Vladeck, *supra* note 33; *see also* Matthew J. Tokson, *The Content/Envelope Distinction in Internet Law*, 50 WM. & MARY L. REV. 2105, 2114 (2009) ("[A]n emerging body of case law suggests that the content/noncontent distinction is crucial in determining whether Internet communications are protected by the Fourth Amendment.").

trine, the searching algorithm will have access to a dataset roughly analogous to that of the drug-sniffing dog.

The second feature pertains to the algorithm: it would have to be programmed to recognize only patterns indicative of illegal activity. This feature sounds deceptively simple. Ordinarily a regression will take certain inputs (locations, purchases, patterns of communication) and turn them into a probabilistic output. To simplify to near the point of absurdity, the algorithm might render a result of "p(terrorist) = 0.9." However, given the enormous complexity of a database that would compile all data available to the government, a sophisticated algorithm would have to create new, intermediate inputs after analyzing the initial variables.[117] Imagine, for example, that there is a pattern of activity that adherents of a certain right-wing group tend to follow prior to committing acts of violence, but such a pattern can also be displayed innocuously by nonmembers; a sophisticated algorithm might be designed to identify both the pattern of activity and the group membership, and flag only those individuals who fit both criteria. Or perhaps sexual orientation could help distinguish between patrons of a given brothel and shoppers at the grocery store on the ground floor. Such intermediate inputs raise questions that the drug-sniffing dog does not pose (imagine, by way of comparison, the drug-sniffing dog thinking to itself, "well that kind of smells like cocaine, *and* he's Colombian, so . . ."). Yet concerns over impermissibly biased analyses can be partially answered by the third and fourth features: that is, so long as the intermediate inputs are not revealed to a human until there has been a highly reliable indication of probable cause for illegal activity, the evaluative process of the algorithm is less significant.

The third feature is that human interaction occurs only at the point of probable cause. Kerr has argued that a computer's analysis of private information is irrelevant to the Fourth Amendment; a Fourth Amendment search should be found to occur only at the moment that a human interacts with private information.[118] Courts could apply the dog-sniffing doctrine to demand a roughly similar process, so that first the algorithm "barks" to indicate probable cause, and then the human search of the information that aroused suspicion follows. For the purposes of this Note, it is assumed that a magistrate would be interposed between the algorithm and the human search[119]: if an individual met

---

117 Taipale refers to this process as "clustering." *See* Taipale, *supra* note 27, at 28–31.

118 Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 554 (2005) ("Applying [*United States v.*] *Karo* and *Kyllo* to computers strongly suggests that a search occurs when digital information is exposed to human observation, not when it is copied from a hard drive.").

119 Whether a warrant should be required is a close question. In the dog context there is no warrant requirement; the search can proceed immediately once the dog indicates probable cause.

the requisite threshold for probable cause, the program could give an anonymized summary of the incriminating information to an agent, who could then seek a warrant for more thorough exploration of the collected data (and potentially other sources of data not accessible to the algorithm, such as email content).[120]

The final critical feature would be the false-positive rate. The appropriate false-positive rate is a minefield well outside the scope of this Note, and might raise questions about whether probable cause is variable across classes of crimes (one imagines that courts would be more accepting of a high-percent false-positive rate for terrorism than for illegal downloading of copyrighted films). It is worth noting that the consequences of a false positive might be regarded as significantly more detrimental to one's privacy in the data-mining context than in the dog-sniffing context. Yet, though a search of one's luggage or car does not yield the same depth of information as a comprehensive examination of one's metadata, the Court has recognized the privacy interests connected to the former as worthy of protection.[121] It is thus not too great a stretch to require the algorithm to provide evidence of reliability sufficient to parallel the faith the Court has placed in the drug-sniffing dog.

C. Is the Crime-Sniffing Algorithm Feasible? Is It Desirable?

The algorithm described here may seem something of a *deus ex machina* solution to the tradeoffs between law enforcement and privacy that data mining poses. Yet such a program is technologically feasible. A drug-sniffing dog need not be perfect; neither data demonstrating significant error rates[122] nor a concrete instance of a dog's misidentification[123] have prevented the use of drug-sniffing dogs in providing probable cause for a more intrusive search. Rather, the Court has made clear that a dog's alert must merely lead a reasonable

---

The Court has excused the warrant requirement in that context due to the exigent circumstances presented by mobile cargo that has been only briefly detained, *see* United States v. Place, 462 U.S. 696, 701–02 (1983), but such an exception would not reasonably apply in the data-mining context (the suspect is hardly likely to begin furiously destroying evidence once a database alerts a government agent to a possible crime).

120 *See* Taipale, *supra* note 27, at 74–81 (discussing technological features that could be built into both databases and algorithm code in order to address privacy concerns).

121 *See* Bond v. United States, 529 U.S. 334, 336 (2000) ("A traveler's personal luggage is clearly an 'effect' protected by the [Fourth] Amendment.").

122 *See* Illinois v. Caballes, 543 U.S. 405, 409 (2005) (rejecting the contention, put forth by both the respondent and the dissent, *id.* at 411–13 (Souter, J., dissenting), that the existence of false positives undermines the basis for the dog sniff's constitutionality).

123 *See* Florida v. Harris, 133 S. Ct. 1050 (2013) (allowing evidence to be introduced from a search based upon an erroneous canine alert).

person to conclude that a search would reveal evidence of wrongdoing.[124]

An automated algorithm can meet this threshold. As Taipale has noted, "[t]he use of probabilistic models developed through data mining can substantially improve human decision-making in some contexts."[125] A crude data-mining program could simply query a database for individuals who match a set of criteria that have been deemed sufficiently indicative of wrongful activity. A sophisticated database could return a list of persons who had purchased certain chemical compounds, recently participated in certain kinds of international financial transactions, and visited identified extremist websites. Comparable mechanical filters are already in place in government communications infrastructure,[126] and there is little reason to think that the NSA is less capable of designing criteria supplying probable cause than Target is of designing criteria suggesting pregnancy.

That a cleverly designed data-mining program might be able to pass constitutional muster is not necessarily a ringing endorsement. The interposition of the algorithm between the dataset and human eyes does not fundamentally alter either of the two core privacy concerns raised by data mining: aggregation of data and investigation without particularized suspicion.[127] The availability of massive amounts of data to the government, subject only to the promise that nobody will look at one's intimate secrets until the algorithm lets them, will inevitably set off alarm bells for those inclined to be suspicious of government. While such concerns are not unreasonable, the alternatives — either not collecting the data or compartmentalizing them in order to inhibit the construction of detailed, intimate portraits — involve largely forfeiting the potential law enforcement gains from data mining. Abandoning pattern-based data mining as a law enforcement tool is appropriately within the scope of public debate, but this Note argues that it is hardly a constitutional necessity.

As for investigation without particularized suspicion, to some extent the algorithm does not change the fact that everyone is subjected to a form of surveillance. Yet the existence of drug-sniffing dogs in Penn Station does not seem to alarm civil libertarians. The critical features of such dogs, as identified by the Court, are that they can pro-

---

124 *Id.* at 1058.

125 Taipale, *supra* note 27, at 33 n.118.

126 *See* JACK GOLDSMITH, BROOKINGS INST., THE CYBERTHREAT, GOVERNMENT NETWORK OPERATIONS, AND THE FOURTH AMENDMENT 3–5 (2010), http://www.brookings.edu/~/media/research/files/papers/2010/12/08%204th%20amendment%20goldsmith/1208_4th_amendment_goldsmith.pdf [http://perma.cc/JGH3-9GAA] (describing the EINSTEIN program's filtering of electronic communications with designated malicious signatures).

127 *See supra* Part I, pp. 693–96.

cess only information about whether one is engaging in an illegal activity, and that they do so without intruding into what one has not already exposed. If an algorithm, with sufficient oversight, could become as innocuous to the public as the dogs — recognizing that HAL and Lassie are not exactly neck and neck in vying for the public's affection — there is no reason why we could not come to regard such light surveillance of our public activities as ordinary.

V. CONCLUSION

"In Algorithm We Trust" is not a slogan likely to catch fire among civil libertarians any time soon. Yet courts must navigate the tradeoff between data mining's boost to law enforcement efficiency and efficacy on the one hand, and its threat to privacy on the other hand. Interposing a carefully designed and regulated machine between the data and the human observer offers the potential to capture significant benefits while satisfying a number of both constitutional and privacy concerns. Close review of the Court's dog-sniffing cases provides four principles by which pattern-based data mining should be regulated: analysis only of legally obtained information; exclusive focus on detecting illegal activity; no human observation without prior probable cause; and low error rates. Even if courts do not apply these features precisely — there are *some* relevant differences between canines and computers — this Note at a minimum demonstrates that dog-sniffing cases provide a template for judicial oversight of pattern-based data mining that involves neither complete dereliction of constitutional oversight nor free-form weighing of costs and benefits, as in the context of the Fourth Amendment's "special needs" doctrine.[128] Rather, courts should carefully scrutinize the dataset, the method and sequence of analysis, and the reliability of the results.

---

[128] The Court has identified certain "exceptional circumstances in which special needs, beyond the normal need for law enforcement, make the warrant and probable-cause requirement impracticable," and in which courts should conduct a "balancing of interests." New Jersey v. T.L.O., 469 U.S. 325, 351 (1985) (Blackmun, J., concurring in the judgment).