ESSAY

THE FOURTH PARTY DOCTRINE: REGULATING BIG DATA WITH AN INFERENCE-BASED APPROACH

Ishan Kumar†

Fourth Amendment law is inadequate to handle the privacy intrusions that are becoming part and parcel of the role Big Data plays in government investigations. The courts view data privacy ontologically through the lens of analog privacy intrusions—that is, they attempt to regulate government access to particular kinds of data, setting standards for what constitutes a search based on the nature of the data requisitioned. However, this approach leaves courts unable to regulate government's use of state-of-the-art technologies which reveal individuals' private lives by drawing inferences from one or more large datasets. These inferences are marketed to government officials, who use them (without a warrant) to gain deeper insights about individuals than those that are provided by the types of information currently protected by the Fourth Amendment. If the current regime persists, the courts could take decades to address inference-driven government investigation. By then, it will have rendered many of the Fourth Amendment's protections moot.

In this Essay, I propose that courts reorient themselves to more effectively monitor Big Data-government collaborations by using: (1) an inference-facing principle to assess algorithmic privacy intrusions from a sounder technological and business perspective, and (2) a fourth party doctrine which applies the existing legal standard for the third party doctrine in Carpenter to regulate the government's ability to deploy Big Data tools and the data science companies which manufacture them. This proposal promises to make Fourth Amendment enforcement adaptive and forward-

† JD/PhD Candidate, Stanford Law School and Stanford School of Medicine. Advised by Professor Robert Weisberg, Stanford Law School.

facing and does not require the creation of new law. Instead, I draw upon fundamental privacy principles articulated by the Supreme Court in Carpenter *to establish a framework for judges to apply and rapidly reshape American privacy law to comport with societal norms. Our proposal has implications for the civil sphere as well; it offers an avenue for privacy regulators to demarcate clear and reasonable rules of conduct for Big Data.*

INTRODUCTION

The business of third-party collection and analysis of our data, "Big Data," is now part of the United States' social fabric. Almost all Americans use digital services or interact with analog technology paired with digital analytics. For example, a home's electricity consumption patterns might be analyzed by a data company to help utility companies better allocate resources to a specific part of a grid at a specific time of day.

Or, a digital watch transmits information on fluctuations in heart rate, location, and device usage to a data company in order to determine how to improve clients' products or targeted advertising. Consumers are now aware that the data they generate in the course of their daily routine is a commodity used by companies which collect, analyze, and apply data (an industry collectively called "Big Data") to improve all sectors of our economy, such as advertising, health care, manufacturing, and software services. The ubiquity of Big Data means data companies act like third parties when they directly provide services to consumers, in the sense that phone companies and the like do.

Big Data also enables data companies to function as fourth parties in certain circumstances. A data company is a fourth party when a third party such as a power company or a telephone company feeds it granular customer-derived data, like the individual power consumption of a set of customers in a defined region. The data company relies upon this information to develop better inferences and services for its own customers or internal processes (e.g., trading strategy in a specialized investment fund). It does not directly interact with the consumers who generated the data it analyzes and is thus a fourth party.

Some entities can simultaneously be third parties (in that they interact with consumers) and also fourth parties (in that they have access to data from other third parties who are directly privy to consumer data). Such an entity is an "omni-party." Many big tech companies have quietly become omni-parties. For instance, Google obtained access to a massive trove of credit card transaction information through a deal with Mastercard and attempted to gain similar access to data from other card payment companies—an effort only reported in 2018.[1] Such information would allow them to see if their targeted advertising was actually affecting user purchase behavior, and thus justify charging their ad-purchasing customers higher fees if their ad services proved to be truly effective. This is an example of fourth-party behavior by an entity traditionally thought of as a third party, and why Google is an omni-party. Such partnerships are not the only way

---

[1] Mark Bergen & Jennifer Surane, *Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales*, BLOOMBERG (Aug. 30, 2018), https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales [https://perma.cc/423W-HJRH].

these parties comingle. Recently, McDonald's acquired a data science company specializing in leveraging consumer behavior analysis for $300 million.[2] At this point, data science applications are so ubiquitous that most third parties are actually behaving like omni-parties in some sense—but not in a fashion that is easily discernible to the consumer (e.g., McDonald's).

The legal and technical mechanisms by which third parties, like Mastercard, transfer their data to fourth parties (or omni-parties like Google) are mysterious. One possibility is that the two entities form a clear contractual arrangement in which the fourth party pays the third party to access its database. At the other end of the spectrum, they might have a parasitic and nonconsensual relationship in which a fourth party employs software to scrape information from a third party's website or database. An extreme (but hardly outlandish) variation of this scenario is that the fourth party scrapes varying amounts of information from numerous third-party websites, all without the consent or knowledge of the third parties. In the middle of the spectrum, a hybrid of these three scenarios might exist: some companies have preexisting data sharing agreements; at least one of them is engaging in nonconsensual data collection from others within the agreement; and at least one of them is collecting data from dozens of unaffiliated third-party companies. As with many commercial agreements, the confidentiality or passive secrecy (e.g., the unannounced revenue-sharing agreement) of deals such as the Google-Mastercard deal, make a proper analysis of the relationship between fourth parties and third parties difficult to accomplish. Compounding the issue is the absence of a clear legal regime (or corresponding mechanism for enforcement) governing the noncontractual relationships described above.

The opacity of the relationship between fourth/omni-parties and third parties means the general public often has no idea where their personal data is ultimately going, and who will handle it once it has been relinquished to a third party. This enables entities like Google, and even the government, to access troves of personal data undetected until they choose to reveal the existence of a deal, or an intrepid investigative

---

[2] Heather Haddon & Diana Mattioli, *McDonald's Buys Israeli Digital Startup Dynamic Yield*, WALL ST. J. (Mar. 25, 2019), https://www.wsj.com/articles/mcdonalds-nears-deal-to-buy-israeli-digital-startup-dynamic-yield-11553552124 [https://perma.cc/GQH7-MWU8].

journalist discovers a particularly egregious example. Worse, the value of such data (and the sharpening of incentives to maintain secrecy) is compounded by its proprietary nature. Omni-parties compete ferociously for rare datasets which can give them an edge over competition. Unrestrained by incentives to the contrary or regulation, corporations readily deploy their data if first-mover advantage is at stake in a market which can be disrupted by algorithms trained on these datasets. Even genomic data, the most personal of all information, is now bartered by firms who seek to develop the first predictive models for a myriad of health-related purposes.[3] Today, in the omni-parties' cost-benefit analyses, the risk of exploiting personal data is almost always miniscule in comparison to the rewards to be reaped by becoming the next Uber, Google, or Salesforce.

Governments are incentivized to behave in the same way. At the international level, the relative strength of data privacy rights in the United States is widely considered to be one of the factors which might place the country at a disadvantage to China's AI development.[4] However, there are workarounds which compensate for this apparent disparity. Entities like Palantir enable the United States to engage in vast data-mining efforts without the same checks to which agencies like the NSA are subject. Although this is old news in the national-security context, Americans must now contend with similar technologies being deployed in every aspect of their lives. Algorithms which operate in the background collect ever-increasing amounts of data from users' interactions with the products they power (e.g., voice-activated digital assistants, internet activity tracking, and cars' on-board computers). These datasets are liquid and difficult to trace; hence journalists' comparison of data to oil is apt.[5] But if government officials' use of a fourth-party algorithm that was trained on

---

[3] *See, e.g.*, GSK Press Release, *GSK And 23andMe Sign Agreement To Leverage Genetic Insights for the Development of Novel Medicines*, GLAXOSMITHKLINE (July 25, 2018), https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/ [https://perma.cc/3PG9-W252].

[4] *See, e.g.*, *The Algorithm Kingdom: China May Match or Beat America in AI,* ECONOMIST (July 15, 2017), https://www.economist.com/business/2017/07/15/china-may-match-or-beat-america-in-ai [https://perma.cc/7B5B-ZRP5].

[5] *See, e.g.*, *Regulating the Internet Giants: The World's Most Valuable Resource is No Longer Oil, But Data*, ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [https://perma.cc/3AWK-RG59].

such data yields intensely personal information about an individual being investigated, that individual currently has no legal protections against such a violation. There might be specific fourth-party algorithms which require a warrant for their use in investigations, but these are highly individualized cases that apply to a tiny fraction of the many algorithm-driven services that are readily available to government officials. The seemingly uncontrollable nature of the interactions between omni-parties, fourth parties, and government present a unique, unforeseen threat to individual privacy in both a Fourth Amendment and a wider privacy rights context.

Government representatives, by coordinating with fourth parties and omni-parties, can peer into the most intimate parts of individuals' lives. They can glean information which is one or two leaps of logic away from the nature of the source data, with the effect that users cannot easily predict what will be learned about them from their behavior. For example, if Google cooperated fully with the government on some issue after its deal with Mastercard, the government might not only have access to transaction history or Google ad analytics to aid its investigations, but also the tools Google might have built to link its normal practice of targeting ads to specific profiles to its trove of Mastercard data.

Generally speaking, the "tools" of Big Data can be boiled down to a data mining pipeline. Data, in any form, is collected and cleaned before being fed into algorithmic models (nowadays, these are frequently neural nets or other deep learning models) which produce inferences from the data that denote the probability of a given outcome. The better the input data, the more incisive and expansive such observations will be. "Better" data means the comprehensiveness (level of detail) and scope (availability of data across time and other domain-specific criteria) of the data available on a particular phenomenon being studied. This is used to improve the models in preparation for their application to real-world problems. If the models are effective and they have been rigorously tested, they will be able to assess new input data in light of what they have learned from the empirical data they were trained on and provide predictions which supplement human decision making. This workflow is why data collection is so important that it is now referred to as the "oil" of Silicon Valley.

In this three-part pipeline (data collection/input, algorithm/model development, and inference generation and

application), regulators' and privacy advocates' attention has been primarily focused on part one, specifically how data is collected from consumers and whether it constitutes a direct violation of privacy rights. This approach failed completely in the last decade, because the development of methods data companies can use to collect data outpaced regulatory action. Also, the incentive for third parties like Mastercard and (more recently) 23andMe to monetize their data stores via licensing or sale grew rapidly, too. Part two of the pipeline is nigh-impregnable to regulators due to the protection of such models under trade secret and copyright law (and, to a more limited extent, patent law). Even the creation of a regulatory body to curate (via review and enforcement of policy) such models and algorithms on a confidential basis is unrealistic given the brutal competitive environment in the technology sector, the paucity of expert talent to staff such a body, and the legal gray zone emerging technologies rely on in order to gain traction (e.g., the well-known Uber/Napster stories). Even maintaining confidentiality in such a process will be a mammoth challenge for such an approach, given the sheer value and liquidity of the digital material being vetted. Part three—the inferences generated by algorithms—is free of the problems which plague the rest of the pipeline, and thus it is an appealing target for the development of a regulatory regime for modern privacy rights: if you can't control what goes in, you might as well control what comes out. Courts have not begun to view the data pipeline from this perspective in their Fourth Amendment jurisprudence. Meanwhile, regulators are still focused on part one of the pipeline—and they are now leaning toward compensating for its intrinsic "slipperiness" with the sheer scale and scope of potential regulation. The GDPR and the massive levies it continues to exact from tech companies inspired them and privacy advocacy groups to consider adopting a similar regime in the civil context.[6]

The courts lag behind the regulators, and the Supreme Court still considers Fourth Amendment data privacy cases in terms of controlling the collection of individual types of data. Yet the sheer diversity of information types employed by omni-

---

[6] *See generally* David Meyer, *In the Wake of GDPR, Will the U.S. Embrace Data Privacy?*, FORTUNE (Nov. 29, 2018), http://fortune.com/2018/11/29/federal-data-privacy-law/ [https://perma.cc/YL7A-DTW2] (explaining how the 2018 California Consumer Privacy Act, modeled after the GDPR, spurred discussion of passing a federal data protection regime).

parties to generate useful inferences means that the Court's traditionally ontological Fourth Amendment treatment of data privacy will not curtail the risk omni-parties present to individual privacy. This practice is equivalent to using a finger to stopper a leaking dam. If it persists, the Court will need to issue multiple rulings on what constitutes a search in the following fields to address even the Google-Mastercard example: ad targeting, personal user profiles, physical and digital card transaction history owned by a noncard payments entity, and the process linking "anonymized" transactions to the profiles which were targeted by Google's ad services. To do so would take the courts decades.

Effectively unfettered until then, government and omni-parties can cooperate to achieve goals which individually appear to be prosocial, such as the apprehension of the Golden State Killer, but in sum profoundly erode a fundamental pillar of modern privacy—a de facto firewall between omni-parties/fourth parties and government which should only be breached after some judicial consideration. Apple's experience with the FBI's demands to unlock an iPhone is a poignant example, while Palantir's Gotham and Metropolis services to government and private entities are at the other end of the extreme—here the fourth party is catering to government needs.[7] The omni-party's perch atop the personal data hierarchy and its increasingly porous relationship with government threatens to swallow privacy rights whole.

This situation has historical analogs. The third party doctrine, by declaring that users could passively surrender their privacy rights, empowered the government to access specific information to aid its investigations. However, subsequent exceptions to the third party doctrine (the Wiretap Act and Electronic Communications Act) tried to reinforce the Fourth Amendment in technological areas which turned out to be fundamental to normal social functions, such as telephone communications.[8] In passing these laws, Congress acknowledged that the third party doctrine should not deprive individuals of Fourth Amendment protections for information transferred in the course of basic daily activities, like phone

---

[7] *See* Ellen Nakashima, *Apple Vows to Resist FBI Demand to Crack iPhone Linked to San Bernardino Attack*, WASH. POST (Feb. 17, 2016), https://www.washingtonpost.com/world/national-security/us-wants-apple-to-help-unlock-iphone-used-by-san-bernardino-shooter/2016/02/16/69b903ee-d4d9-11e5-9823-02b905009f99_story.html [https://perma.cc/YH7K-XXYH].

[8] *See* 18 U.S.C. § 2511 (2012); 18 U.S.C. § 2510 (2012).

calls and emails. Today, in the realm of data privacy, the omni-parties have left us in a post-*Katz*, pre-Wiretap Act world, as Congress has not passed legislation to reinforce the Fourth Amendment's application to personal data—even though individuals can relinquish their data privacy interests under the extant third party doctrine. The current political situation makes such an event unlikely in the near future.

Because the Court extends Fourth Amendment protections for one technology class (or data source/type) at a time, it is always behind the curve for even comparatively limited and dated technologies, like GPS.[9] To make Big Data subject to Fourth Amendment protections, the Court would have to rule upon the hundreds, or perhaps thousands, of data sources omni-parties use in generating useful insights for their customers. This is clearly impractical. To paraphrase *Katz*, the Fourth Amendment should protect people, not databases. A new approach is needed to curtail the third party doctrine's treatment of omni-parties, but not one which requires a wholesale reimagination of privacy in the law. I extract a model for this "Fourth Party Doctrine" from *Carpenter* below.

I
THE PROPOSAL—TARGETING TECHNOLOGIES WHICH CONSTITUTE AN "UPPER BOUND" ON INTRUSIVE INFERENCES ACCESSIBLE WITHOUT A WARRANT

I propose that Fourth Amendment law shift from the futile task of analyzing individual data sources for potential privacy violations and instead focus on determining whether the inferences and insights omni-parties draw from part three of the data pipeline intrude upon privacy norms as defined in *Carpenter*. For example, if Google uses purchase history and geolocation history to infer details about a user (within some margin of error) that the user has tried to hide, and packages such a technology into a warrantless software service used by law enforcement for investigative purposes, it might cross a line set by the Court. This standard can be government-facing, too. Law enforcement officials might be required to obtain a warrant for their use of omni-parties' services if the information they desire derives from inferences which

---

9 *See generally* Florida v. Jardines, 569 U.S. 1 (2013) (holding that using a trained dog to sniff drugs on a property is a search); United States v. Jones, 565 U.S. 400 (2012) (holding that use of a GPS device to track a car is a search); Carpenter v. United States, 138 S. Ct. 2206 (2018) (holding that accessing cell site location data is a search).

penetrate the four spheres of data subject to extra protection defined under *Carpenter*.

Fortunately, this model does not require jettisoning decades of law—the Court already constructed a test in *Carpenter* which needs only to be reaffirmed and explained in this new inference-facing context. This fourth party doctrine is merely a reorientation of existing Fourth Amendment jurisprudence to face a different aspect of the data industry.

The *Carpenter* opinions, in the context of cell-site location information (CSLI), created a multipart test for whether a type of data collection is too intrusive for the government to access without a warrant. But if we imagine that the majority's reasoning is broadened beyond CSLI or data sources generally, the test indicates that there is a search when the information sought is "deeply revealing," has a certain "depth" and "breadth," exhibits "comprehensive reach," and its collection is "inescapable and automatic."[10] Justice Kennedy's dissent helpfully reframes some of these general terms: "intimacy, comprehensiveness, expense, retrospectivity, and voluntariness."[11] In addition, some types of information seem subject to more protection than others: data which can reveal familial, professional, political, and religious information (FPPRI).[12] These concrete data types should be the only inferences which are subject to a broadened, inference-facing interpretation of *Carpenter*, because they are not as open to interpretation by lower courts.

The *Carpenter* factors should provide lower courts with a touchstone for their analysis of the variety of situations they encounter. The extra protection afforded to FPPRI will ensure a common base for lower courts' reasoning and should be applicable to all the inferences and insights omni-parties generate. Under this regime, the government would need a warrant to access FPPRI information which is deeply revealing, broad and comprehensive in scope, and is derived from normal, daily activities.

The focus on FPPRI does not significantly curtail the protection of other kinds of data, and it is specific enough that its scope is moderate. This is because, in order to produce inferences which touch upon FPPRI, data companies often use datasets which are a composite of several data types that might

---

10 *Carpenter*, 138 S. Ct. at 2217.

11 *Id.* at 2234 (Kennedy, J., dissenting).

12 *Id.*

have nothing to do with FPPRI on their face. For example, a pregnant woman might search for a larger car and a new insurance policy. Both of these are not FPPRI. But the fourth party can use this information to conclude, with a certain level of confidence, that the searcher who combines two such searches is pregnant—one of the most sensitive types of familial information. If a judge reviewed the requisition of such data, her analysis would necessarily prevent the source/input data (in our example, the search for a new car and insurance policy) from being applied to divine FPPRI. Thus, the warrant requirement could not be easily circumvented by manipulating the datasets used to produce these inferences. The methods that can currently be used to evade existing privacy protections are varied, and include swapping out one kind of protected data for an equally predictive unprotected data type, using novel combinations of unprotected datasets to produce FPPRI, and even applying a new technique to generate functionally equivalent data to recapitulate the training value of sensitive datasets.[13] The approach proposed here will apply to all of these methods and can be extended to others which are in a similar vein.

This approach is also a simple, effective tool to regulate the government's access to the fruits of neural nets and machine learning. For these technologies, the only discernible part of the three-part pipeline which can reasonably be regulated from a technical perspective is the inference. To become effective, these technologies need free access to training data for experimentation. Their reasoning, the second part of the pipeline, is well known to be a black box. However, their inferences about input data are readily obtainable and are often much more practically relevant than the nature of the input data (but not necessarily easy to understand if one does not know the data which was used to train the machine). In fact, courts are uniquely prepared to resolve ambiguity about the extent to which such inferences reach FPPRI—such issues are similar to what courts decide regularly, and no in-depth knowledge of data science (or expensive expert testimony) is

---

[13] In the civil realm, foreign countries are enacting sweeping changes to part one of the pipeline, in recognition of the kinds of tactics data companies use. For instance, Germany recently restricted Facebook from combining Facebook activity data with other third-party website activity information. Users now have the power to opt out of such practices. *See* Natasha Singer, *Germany Restricts Facebook's Data Gathering*, N.Y. TIMES (Feb. 7, 2019), https://www.nytimes.com/2019/02/07/technology/germany-facebook-data.html [https://perma.cc/ACW8-FFWK].

needed. This proposal provides a clear benchmark for the "upper bound" of technologies which can be used to intrude upon FPPRI and will prevent attempts by the government to use the services of contractors and existing fourth parties to enhance its warrantless search capabilities. However, the inference-facing approach is powerful enough to block even the use of everyday technologies by law enforcement—necessitating an articulation of how the fourth party doctrine applies (or does not apply) to the "lower bound" of technologies.

II
THE "LOWER BOUND"—HOW DOES THIS APPROACH APPLY TO REGULATE GOVERNMENT USE OF CONSUMER-FACING TECHNOLOGIES?

This proposal's absolute upper bound would encompass the most intentional, targeted intrusions into personal privacy—e.g., government officials using software designed to enhance their surveillance capabilities.[14] This is a high-class problem, touching upon national security concerns and the like. However, the problem more likely to affect government and citizens on an everyday basis is that FPPRI is broad enough to potentially prohibit the use of things like a Google search as well. For example, googling a suspect's name will yield information such as accounts on Facebook and LinkedIn, as well as phone numbers, election donation history, real estate, and school information.[15] Much of this data is not voluntarily turned over to the third-party sites which mine and sell access to it. The data is scattered across the internet, but Google's search algorithm is designed to collect it all in one search page for the benefit of a user. So, even a quotidian Google search would require a warrant under the upper bound of this proposal because the search engine's inferences (based on an amorphous weighting of others' search patterns for the same topic and yours) are furnishing the user with suspect-relevant FPPRI.

Obviously, warrant-granting at such a scale is untenable—we should not fetter law enforcement to the extent that routine,

---

14 "Upper" here denotes software which is not ubiquitously available to the public and is expressly designed to infer FPPRI. "Lower" denotes software which is free, very accessible, and can potentially touch upon FPPRI depending on user requests.

15 Given that "FPPRI" includes professional information, gaining access to private LinkedIn profile information and similar types of career data could be prohibited as well.

cursory information-gathering about people becomes onerous. In an age where citizens regularly comb through others' online profiles before in-person interactions, one's expectation of privacy from the government's scrutiny at such a low resolution is not defensible. Thus, a simple solution is to rely on the third party doctrine to establish a lower bound exempting such types of inferences. Depending on how it is applied, an appropriate benchmark could be: if one's information is available to the government via websites found through a generally available search engine, it is exempt from privacy protections.

Using the third party doctrine to solve this lower-bound issue in contexts other than internet searches is not necessarily advisable. The lower-bound issue is broader than just internet searches—it extends to much of our mundane use of consumer-facing technology. For example, how would one apply the third party doctrine in these now-frequent circumstances: law enforcement's use of fake Facebook accounts to surveil suspects, purchase of paywalled FPPRI data, and use of the technology on sites like Ancestry.com to find matches for suspects' DNA? These are all consumer-facing technologies that can readily yield valuable FPPRI to the government, and a straightforward application of the current third party doctrine to each may not be sufficient to establish that a warrant is unnecessary for the undertaking.

A possible solution which can cope with the diversity of such scenarios is for companies to regulate governmental abuse through internal policies akin to the already-extensive infrastructure built to enforce user agreements (e.g., blocking pornographic videos, bot accounts, and the like). They already have the capability to do so. Their incentives are strong, too; users will be warier of furnishing sensitive data if they know that law enforcement has unfettered access. This incentive scales with the complexity and scarcity of that data. Users will likely continue to adhere to social network-type platforms in spite of governmental activity. However, for services like DNA sequencing sites, peer-to-peer transactions, Uber- and Airbnb-type gigs, and others which could provide government officials with insight into users' FPPRI, the prospect of government officials using their technology to surveil anyone on the service could be a serious problem.

While the incentives for self-policing on social networks are weak, it is an indication of how seriously privacy concerns are now threatening data companies that just last year, Facebook

emphasized to the Memphis Police Department that creating fake profiles for law enforcement purposes violates their terms of service.[16] Still, the fundamental issue with companies asserting terms of service violations against the government is enforcement. Technologically, it is not easy to block much of this activity, save for the most egregious kind. Legally, it is not clear what kinds of effective remedies these companies may have against government actors. Without a deterrent effect stemming from a vigorous enforcement mechanism, companies' best self-policing will only succeed in rapidly stanching law enforcement behavior that violates terms of service. Perhaps some kind of private right of action created for companies to obtain injunctive relief for this kind of behavior would be necessary.

Another method by which the fourth party doctrine can be kept from extending too far into everyday algorithmic usage is an objective reasonableness-type benchmark; judges could permit law enforcement to access FPPRI using methods considered generally available to an individual with a layperson's proficiency with technology. This would allow all of the above examples to be conducted without a warrant. However, this "reasonable proficiency" standard will expand the license granted to law enforcement by the third party doctrine to information that is turned over to third parties with what is likely a reasonable expectation of privacy, as in the case of DNA ancestry-matching services.

In lieu of such approaches, perhaps an organic solution to the problem will emerge: judicial discretion. If suspects allege that their Fourth Amendment rights were violated due to an improper, warrantless access of FPPRI through standard services available to the general population, judges might fashion their own standards for their jurisdiction. But this might require decades of jurisprudence to produce a convergent understanding of such a standard that is nationally accepted. It would also prolong the confusion and lack of uniformity the fourth party doctrine proposed here is attempting to solve.

III

---

16 Tim Cushing, *Facebook Tells Cops Its 'Real Name' Policy Applies to Law Enforcement Too*, TECHDIRT (Sept. 27, 2018), https://www.techdirt.com/articles/20180925/18545740715/facebook-tells-cops-real-name-policy-applies-to-law-enforcement-too.shtml [https://perma.cc/5N6W-2ATT].

APPLICATIONS

A series of examples are provided here to illustrate: (1) how the government and omni-parties tend to view the tradeoff between individual privacy and public security, (2) how government officials can leverage the lower bound of technology (as described above) to intrude upon FPPRI with little risk, and (3) how the fourth party doctrine would regulate the upper bound of technologies available to the government, made by companies who tailor their software to service government surveillance efforts.

A. Apple and Lavabit

The positioning of the data company as a sort of information processing plant makes it a very attractive source of information for government investigators, who have for years sought both raw data and inferences to bolster their cases. Perhaps the most dramatic instance of this is the 2016 showdown between Apple and the FBI over the password for an iPhone belonging to a terrorism suspect.[17] Before the situation was defused by the FBI gaining access to the phone independently, Apple faced the prospect of being compelled to develop software to help the FBI circumvent or break through passwords on its product.[18] The conundrum in the case was the trade-off the government demanded: that Apple weaken encryption on its products worldwide in order to help investigators legitimately serve the national interest. Framed differently, the issue was whether a slight-to-moderate (and permanent) decrease in user privacy across hundreds of millions of devices was worth the more efficient investigation of a few dangerous individuals. Unfortunately, with the case rendered moot, we never had the chance to see if such a bargain would be sanctioned by a court.

This bargaining is at the heart of all government information requisition from data companies. While Apple versus FBI did not resolve this question, it had been raised in a more limited fashion in 2013, during the Edward Snowden episode. The FBI sought encryption keys for a secure email service, Lavabit, and the company's founder resisted (the keys

---

17 Katie Benner & Eric Lichtblau, *U.S. Says It Has Unlocked iPhone Without Apple*, N.Y. TIMES (Mar. 28, 2016), https://www.nytimes.com/2016/03/29/technology/apple-iphone-fbi-justice-department-case.html [https://perma.cc/2LMV-MPGY].

18 *Id.*

would have given the government the capability to monitor every user's email communication). In response, a federal district court fined his company $5,000 a day until he complied with their order. Two days later, he capitulated, turned over the keys, and then shut the service down entirely.[19] On appeal in the Fourth Circuit, Lavabit's founder challenged a pen/trap order on the email service and the lower court's contempt finding unsuccessfully.[20] Lavabit's saga was quite influential; it was considered a potential template for the outcome of Apple's dispute with the FBI.[21]

For Apple and Lavabit, yielding even a very narrow technological concession to the government was not worth the risk of the narrow disclosures infringing upon user privacy. However, the courts have been hesitant to wade into the constitutional issues surrounding warrants on technology used to secure user privacy. In the Lavabit case, the district court found that the seizure warrant issued for the encryption keys was "very narrow [and] specific"[22] and the appellate court applied the constitutional avoidance rule to dodge questions raised by Lavabit regarding the constitutionality of the warrant.[23]

These constitutional questions can no longer be neatly sidestepped today. *Carpenter* confronted them to some extent, but it was methodologically very consistent with prior third party doctrine cases (discussed in legal background). It left data companies who are third and fourth parties out of its scope, and also did not address a new phenomenon of a middle ground reached between data companies and law enforcement—seeking inferences from data companies as opposed to raw data (e.g., cell site location information in *Carpenter*). This is not objectionable for data companies, who are happy to avoid disclosing sensitive business information, for users, who are less alarmed by companies granting the government access to services available to customers (partially because the government is not gaining some kind of enhanced

---

19 United States v. Lavabit, LLC, 749 F.3d 276, 284 (4th Cir. 2014).

20 *Id.*

21 John Markoff, Katie Benner & Brian X. Chen, *Apple Encryption Engineers, if Ordered to Unlock iPhone, Might Resist*, N.Y. TIMES (Mar. 17, 2016), https://www.nytimes.com/2016/03/18/technology/apple-encryption-engineers-if-ordered-to-unlock-iphone-might-resist.html [https://perma.cc/6U56-BW95].

22 *Lavabit*, 749 F.3d at 284.

23 *Id.* at 293.

access not otherwise available to private citizens), and for the government, as inferences cut down on internal data analysis and contain whatever information they need up-front, at the same level of quality that customers expect (and avoid the public outcry, as an added bonus). But even though data companies' turning over raw data was (and remains) a major privacy concern for the past decade, we should be much more worried about this middle ground, which is a wolf in sheep's clothing.

Inferences drawn by data companies' algorithms (and now the predictions derived by neural networks) are valuable primarily because they yield more practical information for customers than the raw data itself. Easy, unfettered access to such inferences empowers the government far more than fighting to get a subpoena or warrant for access to a database or the inner workings of an algorithm. They also lose the ability to screen out unrelated information about individuals, and end up analyzing an output that is often a prediction based upon large amounts of training data and is empirically true about a swathe of people unrelated to an investigation (the numbers of unrelated people could range from a single person to the millions and billions). In the Apple and Lavabit cases, the immediate concern was that the government would abuse its newfound ability to circumvent security measures. Here the issue is that plenty of unrelated (and often implicit) information is part and parcel of the inferences requested, making more serious violations of privacy occur more frequently and in a passive manner.

B. Golden State Killer

The recent Golden State Killer (GSK) case, and the wave of DNA search-powered cold case resolutions that followed, epitomize this new phenomenon of inference-based searches. In particular, their frequent flirtations with procedural and substantive (what information is collected) grey zones illustrate the wider problem well. The GSK was a serial killer and rapist who had frustrated law enforcement for decades. In 2018, they identified and apprehended him, but through means which were rife with fundamental privacy issues. As described in the *New York Times*, "the breakthrough in the [Golden State Killer] case came after [investigators] created a fake profile with a sample retrieved in 1980 and pretended to be someone

researching family history."[24] The fake profile was on an unregulated website called GEDmatch, which offers a free genealogical tracing service for people who upload their DNA sequencing data (produced by services like 23andMe). The investigators who uploaded the GSK's crime scene DNA to GEDmatch were using a type of genealogical matching which relies on using partial overlaps between profiles of the aspects of DNA which distinguish individuals from each other, and are largely inherited.[25] The efficacy of such an approach depends upon the number of sequences in a DNA database, up to a certain point. GEDmatch's database is approximately one million users strong, a mere ~0.3% of the US population.[26] Yet a "distant" relative of the GSK had uploaded his information, and it showed up in the results as a partial match after GEDmatch's programs compared the GSK's data to the profiles of a million users.[27] This was not an ironclad prediction; DNA genealogy has patchy accuracy which drops with an increase in genetic distance between family members.[28] Also, the investigators were not being particularly inventive beyond their use of a non-FBI database. The use of familial DNA to find criminal suspects is a notion that dates back at least to 2006.[29] Additionally, some companies directly offer such services to law enforcement.[30]

Of the three parts of a data pipeline that are legally relevant (database, algorithm, inference), two would have been useless to law enforcement in the GSK case even if they had gotten complete access to either. Being able to access GEDmatch's database would be meaningless to investigators, as they would need to write programs of their own to match

---

24 Thomas Fuller, *Golden State Killer Prosecutors Team Up to Find the Suspect*, N.Y. TIMES (Aug. 21, 2018), https://www.nytimes.com/2018/08/21/us/golden-state-killer-trial-sacramento.html [https://perma.cc/7JPD-VRKE].

25 Drake Bennett & Kristin V. Brown, *Your DNA Is Out There. Do You Want Law Enforcement Using It?*, BLOOMBERG BUSINESSWEEK (Oct. 27, 2018), https://www.bloomberg.com/news/features/2018-10-27/your-dna-is-out-there-do-you-want-law-enforcement-using-it [https://perma.cc/F6N8-3U3B].

26 Sarah Zhang, *How a Tiny Website Became the Police's Go-To Genealogy Database*, ATLANTIC (June 1, 2018), https://www.theatlantic.com/science/archive/2018/06/gedmatch-police-genealogy-database/561695/ [https://perma.cc/GFN8-8KPH].

27 *Id.*

28 Bennett & Brown, *supra* note 25.

29 Frederick R. Bieber, Charles H. Brenner & David Lazer, *Finding Criminals Through the DNA of Their Relatives*, 312 SCIENCE 1315, 1315–16 (2006).

30 Bennett & Brown, *supra* note 25.

their sample to any of a million sequences (not a trivial task). Being able to view how the data was matched, without the database associated with the program, would also have been useless. Both might have required a subpoena at the very least. But access to the inference—the result generated by the two together—was simple, exceedingly cheap, and much more informative than other searches would have been.

The issues inherent in applying the fourth party doctrine to the lower bound of technologies persist in the GEDmatch scenario. Would judges require that law enforcement officials go to the courts before making any accounts on such sites? Would they flag fraudulent accounts made for assisting with investigations as being Fourth Amendment violations? The danger is that, left to their own devices, judges in trial courts will be faced with a tough dilemma in which they have to balance protecting poorly defined, eroded rights against the urgent public need to remove individuals like the GSK from society. Indeed, their preference for the latter is borne out in many of the cold cases being solved by law enforcement officials through analogous DNA-matching tactics.[31]

Still, even in these lower bound cases, it is the inference which should be at the crux of judges' attention. It is the most dangerous part of the data company's apparatus, because it can reach far beyond the scope of normal search parameters. An inference is not analogous to a search of a car, or even a search of a phone. Those repositories of information are wholly within the control of their proprietors. The inference can allow investigators to bypass attempts by individuals to sanitize their lives, because no individual can reasonably manipulate the multifarious variables of their daily activity that generate data which is collected and analyzed. The often-opaque outputs of algorithms (e.g., a composite score) are difficult to challenge effectively, and are easy to take at face value. Equally concerning, many times the inferences can be wrong—but there might not be a simple way to check. The technology used in the GSK case was shown by a study to be incorrect 83% of the time.[32] If DNA collected at a 1985 crime scene is uploaded

---

31 An added complexity: many such services might share data with hospital systems and the like. Queries would touch data that might still be covered by patient privacy laws or similar legal regimes.

32 Avi Selk, *The Ingenious and 'Dystopian' DNA Technique Police Used to Hunt the 'Golden State Killer' Suspect*, WASH. POST (Apr. 28, 2018), https://www.washingtonpost.com/news/true-crime/wp/2018/04/27/golden-state-killer-dna-website-gedmatch-was-used-to-identify-joseph-deangelo-as-

today, the investigator risks propagating the harm caused by human error in sample handling and immature forensic science protocols. Before, the DNA would have sat ineffectually in a storage room; now it is being deployed and might mistakenly implicate innocents.

The current burden the third party doctrine places on the user, who ostensibly has no reasonable expectation of privacy for the information he willingly turns over to the third party, is misplaced and unfair when viewed from an inference-centric perspective. How is an ordinary person able to anticipate ex ante what kind of information might be gleaned from the data they turn over? Indeed, even experts would struggle to do this; a data company is unlikely to publicly disclose a cutting-edge approach to finding valuable inferences, so much of the technological landscape is shrouded (take Palantir, for instance). Instead, a special modification of the third party doctrine applied in cases which use inferences as evidence should place the burden on the government and data companies to characterize the intrusiveness of an inference for the benefit of a judge, who can then determine whether use of the inference constitutes a search, and whether it rises to the level of requiring a warrant. Fortunately, we do not need to look far to identify a toolbox for line-drawing purposes; the *Carpenter* majority and concurrences fashioned a convenient test that can be repurposed here.

The aforementioned examples of the potential usefulness of an inference-based understanding of data privacy lie between the upper and lower bounds of the technologies to which the fourth party doctrine will apply. The government's relationship with Palantir and its analogs is an example of the upper bound and demonstrates the fourth party doctrine's efficacy in contending with such technologies.

C. Palantir

Today, a burgeoning industry is catering to government officials' appetite for inferences that reveal, among other things, FPPRI. Law enforcement's use of dedicated portals and technologies provided by omni-parties does not require a warrant even though there is a clear risk of FPPRI being revealed regularly through the kinds of algorithmic pipelines data companies offer the government. These could be as

---

suspect-police-say/?noredirect=on&utm_term=.1cf3f9bc9787 [https://perma.cc/NHK7-MAC3].

sophisticated as Palantir's Gotham program, which was used by the LAPD in Operation Laser (a predictive policing experiment) to generate lists of people who were subject to extra scrutiny.[33] The decision to add people to this list was predicated on Gotham's analysis of subjects' recorded contacts with the department and its assessment of their propensity to engage in criminal behavior.[34]

Palantir now services law enforcement at every level of government. It hopes that its Foundry product will allow law enforcement to use its data mining services without the direct aid of onsite Palantir engineers. Foundry eliminates the standard divide between the omni-party and government, in which a set of services are performed in-house at the omni-party and the government officials can act upon the generated inferences. Now, local, state, and federal law enforcement (not just the NSA or CIA) can engage in unregulated, rudimentary data mining of their own on the populations they police.

How does Palantir plan to mitigate the privacy concerns stemming from granting law enforcement officials the kind of inferences normally generated and analyzed behind closed corporate doors? Its concession to the risk posed by such a massive hoovering up of individuals' FPPRI and other data which qualifies for protection under the *Carpenter* majority's opinion is: (1) a small team lacking the bandwidth to handle Palantir's thousands of clients, and (2) internal "privacy protective technology" that partitions law enforcement's access to data records in a granular fashion mirroring current analog practice.[35] However, these measures guard against only the most extreme forms of Fourth Amendment violations—the illegal access of personal information by the government. The inferences produced by these data remain untrammeled.

Our fourth party doctrine would handle the Palantir Foundry scenario. With such a doctrine in place, a warrant would be required for law enforcements' use of programs which yield FPPRI. Hypothetically, if software used by the NYPD inferred from a suspect's arrest record or prison records that

---

33 Martin Kaste, *How Data Analysis Is Driving Policing*, NPR (June 25, 2018), https://www.npr.org/2018/06/25/622715984/how-data-analysis-is-driving-policing [https://perma.cc/KB4X-RLTU].

34 *Id.*

35 PALANTIR, PALANTIR & LAW ENFORCEMENT: PROTECTING PRIVACY AND CIVIL LIBERTIES 4, https://www.palantir.com/wp-assets/media/capabilities-perspectives/Local-Law-Enforcement-PCL-White-Paper.pdf [https://perma.cc/S5WT-3KWZ].

he is Muslim, and this added factor is weighted heavily by that predictive policing algorithm as being indicative of a high-risk individual, the fourth party doctrine would block the use of the program in this instance, leaving the predictive policing algorithm intact for useful applications that are in compliance with our understanding of the Fourth Amendment. Interestingly, technological solutions could be built to filter these kinds of results before they reach users.

IV
BEYOND FPPRI: APPLYING THE FOURTH PARTY DOCTRINE TO PROTECT OTHER, DEEPLY REVEALING FORMS OF INFORMATION

In *Carpenter*, FPPRI was established as a sort of specially protected class of information, but the Court also set forth two tests that help us apply the fourth party doctrine to non-FPPRI—one in the majority opinion, and another in Justice Kennedy's concurrence. While the latter is possibly intended to be a restatement of the majority's test, several variations in the terminology chosen by the Justice lend themselves to a substantially distinct test. These are explored here, and their relative strengths and optionality they grant to trial courts are assessed through the use of illustrative examples.

A. The Majority's Test

The majority defined the scope of non-FPPRI material which is considered sensitive enough for a warrant to be that information which is "deeply revealing," has a certain "depth" and "breadth," exhibits "comprehensive reach," and its collection is "inescapable and automatic."[36] These are broad terms which lend themselves to heavy interpretation and circumvention by trial courts. At their core is how the majority prizes individuals' consent and ability to protect that which is nebulously regarded as too personal—and it seems they use the "ick factor" to gauge this.

In situations absent considerations of FPPRI, the majority's test will result in divergent outcomes in questions over whether government requisition of data from fourth parties or omni-parties requires a warrant. Take, for example, the recent use of Amazon Rekognition by law enforcement (noted in the news for associated revelations regarding bias in facial recognition software)[37] and FamilyTreeDNA's

---

36 Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018).

37 Cade Metz & Natasha Singer, *A.I. Experts Question Amazon's Facial-*

collaboration with the FBI in sharing genetic data from its consumer tests.[38]

1. *Amazon Rekognition*

Amazon Rekognition is a facial recognition AI software sold by Amazon. It has the ability to sift through surveillance footage and identify suspects based on preexisting mug shot databases. Recent controversies regarding its efficacy and biases aside, it was piloted by the sheriff's department in Washington County, Oregon, and by law enforcement in Orlando, Florida. Amazon wished to pitch Rekognition to the public sector at large—and did so, attempting to gain Immigration and Customs Enforcement's business in late 2018.[39]

For the purposes of our analysis, Rekognition is quite similar to the products Palantir sold the government for predictive policing. It is marketed directly to law enforcement agencies at the local, state, and federal levels and it facilitates warrantless searches of private citizens which don't directly reach FPPRI. Under the majority's test, Rekognition would likely be permissible in some types of applications. The analysis does not change when one considers that a true positive match would result in law enforcement being able to apply all the information they already know about a suspect to the identified individual. This would include the individual's history with law enforcement, and whatever FPPRI information is located in those records. However, as this information is previously known to law enforcement, no expectation of privacy will be violated due to Rekognition's matching function. On the other hand, new information garnered by investigative efforts which are based on a positive match might still be protected.

Rekognition's application in law enforcement contexts

Recognition Technology, N.Y. TIMES (Apr. 3, 2019), https://www.nytimes.com/2019/04/03/technology/amazon-facial-recognition-technology.html [https://perma.cc/BB3H-4MM7].

38 Salvador Hernandez, *One Of The Biggest At-Home DNA Testing Companies Is Working With The FBI*, BUZZFEED NEWS (Jan. 31, 2019), https://www.buzzfeednews.com/article/salvadorhernandez/family-tree-dna-fbi-investigative-genealogy-privacy [https://perma.cc/HL6M-B9BR].

39 Drew Harwell, *Amazon Met With ICE Officials Over Facial-Recognition System That Could Identify Immigrants*, WASH. POST (Oct. 23, 2018), https://www.washingtonpost.com/technology/2018/10/23/amazon-met-with-ice-officials-over-facial-recognition-system-that-could-identify-immigrants/?utm_term=.f87626838700 [https://perma.cc/2YNK-JWAS].

could identify suspects in surveillance footage that is "deeply revealing" about an individual in the way that CSLI is. For instance, by scanning enough footage requisitioned from private and public security systems, law enforcement could build a comprehensive picture of an individual's activities—down to what they eat, when they sleep, where they travel, and with whom they associate. This also goes to Rekognition's "depth" and "breadth." Whether Rekognition's capability constitutes "inescapable reach" is a different matter—and would be the grounds upon which courts would permit warrantless use of the platform in some instances. The aforementioned scenario, in which law enforcement officials construct an all-encompassing model of a suspect's behavior, is heavily contingent upon the practicalities of collecting such a large volume of surveillance footage. While public surveillance footage might be readily available, acquiring private footage at that scale would be onerous. Even finding the locations a suspect visited would be difficult.[40] On the other end, building a database constituted of images of the general public is much more fraught than simply using convicts' mug shots to find matches. But for a drastic—yet entirely feasible—centralization of such footage, the implementation of live feeds, and expansion of its dataset, application of Rekognition will usually be limited to the kinds of footage criminal investigators prize (that of a crime being committed, etc.).

Finally, Rekognition's ability to spot suspects—the inference that is the focus of the fourth party doctrine—is far from perfect.[41] While the collection of video footage and its possible live feed to law enforcement could be construed as being "inescapable and automatic" under the current third party doctrine, the Rekognition program itself does not reach that standard when one considers the majority's conception of the factor in the CSLI context. Cellular signals' receipt by towers is certainly inescapable and automatic, because all cellular phones work in this fashion, but Rekognition's

---

[40] However, live feeds from such cameras, including police body cameras, are available to law enforcement and were used by Washington County's sheriff. *See* Elizabeth Dwoskin, *Amazon Is Selling Facial-Recognition to Law Enforcement—For a Fistful of Dollars*, WASH. POST (May 22, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/05/22/amazon-is-selling-facial-recognition-to-law-enforcement-for-a-fistful-of-dollars/?noredirect=on&utm_term=.07002534846c [https://perma.cc/678T-QHQF].

[41] Metz & Singer, *supra* note 37.

matching is not done extensively through live feeds yet.

Two hypotheticals modifying some of the facts in the Rekognition example illustrate circumstances in which it could meet each of the two components of the "inescapable and automatic" factor and thus have it support the need for a warrant. If its accuracy improved to the point that it could be passively deployed and follow-on human confirmation would be unlikely to catch an error, it might be considered "automatic." Alternatively, if it could scan a truly comprehensive camera network, like that in London or many cities in China, it might be considered "inescapable." Of course, combination of these two hypotheticals would indeed qualify Rekognition to be "inescapable and automatic" to the point that its use in such a fashion would be a violation of the fourth party doctrine.

Taken together, the third party doctrine offers precious little to regulate government's use of services such as Amazon Rekognition. The video footage is already collected, and presumably suspects have already relinquished their right to privacy if the footage is properly acquired by law enforcement. However, the fourth party doctrine will effectively curb those applications of Amazon Rekognition which are moderate-to-extreme in their intrusiveness. Generally, under the fourth party doctrine, the more the algorithm improves, and the better the quantity of data it is fed, the more it deserves judicial scrutiny for potential Fourth Amendment violations. Note that the fourth party doctrine will preserve efficient and speedy warrantless applications of Amazon Rekognition in certain settings. These settings might be limited by the quantity or quality of data furnished to the algorithm, or the subpar accuracy of the algorithm itself—as discussed previously, if the algorithm is inaccurate to the point that human double-checking is necessary, it cannot be considered "inescapable and automatic" under the fourth party doctrine and the majority's intended meaning of this factor. This built-in ability to produce balanced, nuanced outcomes makes the fourth party doctrine ideal for judges to follow when they examine programs which are analogous to Amazon Rekognition.

But how would the fourth party doctrine handle a completely different type of data and algorithmic application—DNA and ancestry?

2. *FamilyTreeDNA*

Early in 2019, the company FamilyTreeDNA announced

that it would allow the FBI to access its database of approximately two million genomic profiles of customers who used its direct-to-consumer genetic testing services.[42] It was the "first time a commercial testing company has voluntarily given law enforcement access to user data."[43] The company described the nature of the collaboration as providing the same sort of access that a normal customer would have—in that the FBI would submit DNA samples to its lab, which would then search for matches within FamilyTreeDNA's database.[44] Effectively, this would amount to the FBI querying FamilyTreeDNA's database to find either exact matches or the suspect's relatives. While privacy advocates will undoubtedly be concerned with the precedent this will set (in theory, access to the genomic profiles of a small proportion of the population will allow law enforcement to match almost any suspect to a potential relative), the fourth party doctrine will be able to block overly intrusive applications of FamilyTreeDNA's services. Importantly, the inference-facing fourth party doctrine should play a role here even if the fourth party voluntarily provides government access, as the database itself does not matter—the government's decision to query the genealogical service would bring it under the Fourth Amendment's purview and raise a fourth party doctrine issue.

Applying the majority's test to FamilyTreeDNA will yield an analysis which is significantly different from that of Amazon Rekognition. First, it could satisfy the "deeply revealing" factor for a search, as exact DNA matches or genealogical matching would reveal some associated information about the profile or account on FamilyTreeDNA. On the other hand, that kind of information on its own is not necessarily as significant as CSLI is—it merely identifies an individual, contact information, and who their genetic relations are.[45] Next, this kind of profile matching technologically mirrors the "depth" and "breadth" of CSLI, in that genomic information is fundamental to an

---

42 Kristin V. Brown, *Major DNA Testing Company Sharing Genetic Data With the FBI*, BLOOMBERG (Feb. 1, 2019), https://www.bloomberg.com/news/articles/2019-02-01/major-dna-testing-company-is-sharing-genetic-data-with-the-fbi [https://perma.cc/Z25K-QV3F].

43 *Id.*

44 *Id.*

45 This is where DNA profile matching should be distinguished from "family" as envisioned by the Court in its definition of FPPRI – presumably, the Court meant it to apply to those of blood relation who maintain interpersonal connections. This is just a subset of all those who could be discovered through a FamilyTreeDNA match.

individual, and the range of potential matches is enough to include all of one's genetic relations. The "comprehensive reach" of this service depends on the size of the database to which the FBI has access. However, as a journalist points out, the FBI can already access GEDmatch, an open access database roughly equivalent to FamilyTreeDNA's database size.[46] Further expansions of government's burgeoning DNA dataset could reach a point that would (as mentioned above) enable government to find a blood relative of any suspect. This scenario would certainly meet the *Carpenter* threshold for "comprehensive reach."

Finally, the current FamilyTreeDNA software is much more "inescapable and automatic" than Amazon Rekognition. At any point, officials can search its database and find hits on suspect DNA, using a reliable and established technology. This means the general public is highly susceptible to inferences from the technology, and the general public cannot curb its use without some significant erasure of not just the FamilyTreeDNA database, but all the existing analogous databases available to the government (the "inescapable" component). The "automatic" component is present here, in that the subject does not have a choice about whether to limit FamilyTreeDNA's ability to match relatives to one's sample—the inference generated by the platform. But this can be provided to users through simple measures that are not enabled on such platforms because they reduce the appeal of the service to those users who are intent on exploiting the database akin to how the FBI uses FamilyTreeDNA. One measure which could result in the service passing muster on this factor is two-party consent for establishing matches.

B. The Concurrence's Test

The concurrence helpfully refines the "too personal" terms of the majority's test into easily quantifiable metrics. Justice Kennedy provides that "intimacy, comprehensiveness, expense, retrospectivity, and voluntariness" are the components of his interpretation of the majority's test.[47] Yet, if the majority's test is too broad, his test is perhaps so narrowly tailored that it will apply to certain technologies' inferences perfectly, but not at all to others. Also, none of these

---

46 Brown, *supra* note 42.

47 Carpenter v. United States, 138 S. Ct. 2206, 2234 (2018) (Kennedy, J., dissenting).

factors provide judges with the sort of common escape hatch that gives leeway for extensive judicial discretion (e.g., the test for granting a preliminary injunction). Here, the above two examples of Amazon Rekognition's and FamilyTreeDNA's collaboration with law enforcement are subjected to the concurrence's test in an effort to demonstrate how outcomes might differ or converge under the two *Carpenter* tests.

1. *Amazon Rekognition*

Under the concurrence's test, Rekognition is highly likely to be constrained by the fourth party doctrine and is an example of technology to which the concurrence's test is well suited to regulate government's access. The "intimacy" prong encompasses FPPRI but is also a stand-in for the aforementioned "ick" factor—the instinctive aversive response of the objectively reasonable person to the prospect of a certain kind of search. Amazon Rekognition could very well produce such a reaction. In its most extreme form, an individual could be tracked as that individual progresses through the basic aspects of life, as the Court found to be the case with CSLI. The comprehensiveness prong is functionally equivalent to the majority's comprehensive reach factor. As for expense, the costs of purchasing access to this software are negligible; they can be as low as six dollars a month for an entire county.[48] Rekognition, by dint of its inherent capability to survey surveillance footage from any time in a suspect's life, fulfills the retrospectivity prong. The voluntariness prong, if understood to be a substitute for the "inescapable and automatic" factor in the majority's opinion, would undergo the same analysis. Voluntariness would also be critical in ensuring that matching faces to databases of mugshots will be the extent of the warrantless use of the technology. Ex-felons have curtailed rights, and the availability of their mugshot to Rekognition will certainly not be a major deprivation when compared to others, like the right to vote. Indeed, this sort of information is already available to law enforcement. In contrast, trying to match suspects to databases comprised of individuals without prior criminal records would be prohibited under an inference-based understanding of the fourth party doctrine—this would be involuntary, inescapable, and automatic.

---

48 Dwoskin, *supra* note 40.

2. *FamilyTreeDNA*

FamilyTreeDNA is intimate in the sense that unfettering law enforcement's access to DNA profile matching capability likely satisfies the "ick" factor, even if it does not qualify as familial information in the sense that the *Carpenter* Court conceived of it. Its comprehensiveness depends on the size of its DNA profile dataset and the underlying accuracy of the software. If the dataset grows enough (or achieves an appropriate size when paired with other third-party datasets) that almost every input will result in a match with a genetic relation, FamilyTreeDNA's collaboration with the FBI will be comprehensive. The costs associated with using this service are not negligible. The government would need to collect, process, and send samples for analysis. At scale, this would become quite costly and time-inefficient. Plus, FamilyTreeDNA and its ilk would have a maximum number of samples they can process in a given time. This is where the concurrence's test deviates significantly from the majority and might result in a court permitting such collaborations under the fourth party doctrine until such a time that they become integrated and efficient. Because individuals involuntarily subject themselves to the possibility of being matched against a genetic relative who is in the FamilyTreeDNA database, the voluntariness factor would weigh against the FBI-FamilyTreeDNA collaboration's warrantless searches.

V
THE VALUE OF THE INFERENCE-FACING APPROACH IN DIRECTLY REGULATING GOVERNMENT'S BIG DATA-DRIVEN INVESTIGATIVE EFFORTS

While government-developed algorithms do not directly fall under the fourth party doctrine, they are still susceptible to the inference-facing approach. Judges, when confronted with privacy disputes over such programs, could use the inference-facing approach to prevent such programs from violating Fourth Amendment rights. It would enable them to set broadly applicable standards for such efforts, while being able to adapt to the unique circumstances which accompany new technological innovations. The need for such a structure is dire. Law enforcement is, in addition to availing itself of the services of fourth parties, developing its own software under a much deeper veil of secrecy than private companies who procure government contracts can sustain.

Examples of recent cutting-edge, in-house law

enforcement technology developed outside of the national security legal regime are just beginning to emerge. In lieu of detailed reporting on such software, we can look to the public comments of high-ranking officials themselves to gain insight into law enforcement-developed algorithms. The SEC's employees speak frequently about their organization's use of big data to detect the myriad suspicious practices under its purview.[49] As Scott Bauguess, then-Deputy Chief Economist, stated in 2017, "the Commission has made recent and rapid advancements with analytic programs that harness the power of big data. They are driving our surveillance programs and allowing innovations in our market risk assessment initiatives."[50] He went on to describe some of the SEC's early successes with using natural language processing of reports to spot candidates for extra scrutiny.[51] Two years later, these programs are undoubtedly more sophisticated and integrated with its day-to-day operations.

According to how they were described in 2017, the reams of bulk data the SEC receives hourly are analyzed with a variety of machine learning techniques in order to draw inferences and predictions about individual behavior. They do this without judicial oversight guiding their decisions to use their software in investigations. In one of their earliest successful applications in 2016, one such program, called Artemis, was used to spot and bring charges against a suspect for insider trading. It "analyze[s] patterns and relationships among multiple traders. . . . The SEC also uses software from privately-held Palantir Technologies, which identifies links between individuals and entities by connecting pieces of information from multiple data sources."[52] While the fourth party doctrine would apply to the Palantir relationship, some additional oversight over Artemis-like programs is needed. In a highly specialized sector like the SEC's, the inference-facing

---

49 Scott Bauguess, Deputy Chief Economist & Deputy Dir. *The Role of Machine Readability in an AI World*, SEC. & EXCHANGE COMM'N (May 3, 2018), https://www.sec.gov/news/speech/speech-bauguess-050318 [https://perma.cc/P927-ZNAJ].

50 Scott Bauguess, *The Role of Big Data, Machine Learning, and AI in Assessing Risks: A Regulatory Perspective*, SEC. & EXCHANGE COMM'N (June 21, 2017), https://www.sec.gov/news/speech/bauguess-big-data-ai [https://perma.cc/QE93-REE3].

51 *Id.*

52 Nate Raymond, *Newest Weapon in U.S. Hunt for Insider Traders Paying Off*, REUTERS (Nov. 1, 2016), https://www.reuters.com/article/us-usa-insidertrading-insight/newest-weapon-in-u-s-hunt-for-insider-traders-paying-off-idUSKBN12W2X4 [https://perma.cc/5EVC-8TW4].

approach can still be applied to regulate such programs, albeit with a different, industry-specific set of standards than the ones used in the *Carpenter* opinions.

VI
APPLICATIONS OF THE INFERENCE-FACING PRINCIPLE IN THE CIVIL SPHERE

If this inference-facing principle is applied to data companies' products when citizens bring suit under a private right of action or a regulatory body exercises its enforcement power, the entire technology industry will be forced to curb the worst excesses of its wanton commoditization of user information. Such an effect and its ramifications deserve further examination and thought by scholars, judges, and academics. New regulatory regimes will likely be required to enact such a shift in current thinking on data privacy, but it is an opportune time to consider such activity in the United States. Regulatory bodies have a precedent for far-ranging data privacy legislation in the GDPR, and they are eager to avoid its well-known shortcomings (which are not mitigated by levying historically large fines on Google intermittently). Rather than place the burden on consumers to individually enforce low-value private claims or attempt to form often futile class action/mass tort cases, the inference-facing approach will squarely leave the judiciary in charge of determining to what extent data science's latest predictions should be deployed without triggering regulatory enforcement. It will also provide industry with clear guideposts for preferentially developing the kinds of technologies which enhance the nonintimate aspects of our lives, while treading more carefully with those technologies that delve too deep.

Potentially, a new kind of government clearinghouse for technology can be developed. Unlike those previously proposed, it would not unduly impede algorithm development; nor would it compel data companies to relinquish trade secrets. The requirement would be simple: all data companies must inform the agency of all of the products they offer their customers. The agency would then assess the likelihood and severity of privacy violations and provide standard risk metrics for the benefit of judges and regulators.

CONCLUSION

Twenty years ago, the founder of Sun Microsystems told

the populace, "You have no privacy. Get over it."[53] Many others have since grandiosely declared that privacy is dead in the twenty-first century. This Essay shows that, to the contrary, holding government and data companies accountable to traditional notions of inviolable personal information is not a fool's errand, doomed to either overregulate data collection or underregulate its applications. Nor must regulators reduce themselves to competing to impose ever-larger fines on innovative companies. Rather, our judiciary should begin to think about data privacy the way data companies and their clients do, and fashion a regime flexible enough to accommodate future technological development.

This monumental task demands that judges shift to a forward-facing posture in their evaluation of algorithmic privacy intrusions. The fourth party doctrine, by directly adapting the *Carpenter* rationale for its purposes, incorporates the Court's most current reasoning to establish a starting point for lower courts to begin generating precedent and guidance for government and data companies to follow. The sphere of privacy established by judges' efforts to enact the fourth party doctrine will impose certitude in data regulation and provide a forum for new technologies to be evaluated impartially. Data companies will value regulatory certitude and relief from overly expansive regulations, but they will relinquish their current ability to market intrusive inferences at will. The government's actions, in exploiting the fruits of Big Data, will finally be subject to judicial review before an intrusion occurs—an established process akin to obtaining a warrant for a car or a house. Citizens benefit from having a public accounting of why certain technologies are considered exceptionally intrusive. They can also use this information to agitate for more robust reforms.

The fourth party doctrine, and the inference-facing principle upon which it relies, will modernize the Fourth Amendment, preserve the societal benefits of data science, and restore a sense of basic privacy expectations to the nation. Crucially, it will force the legal system to break from its shameful behavioral pattern. First, there is its reluctance, born out of fear and ignorance, to grapple with cutting-edge technologies. Next, there is a retreat toward regulating only

---

53 Megan Mcardle, *Take My Privacy, Please! A Defense of Google*, ATLANTIC (Mar. 8, 2012), https://www.theatlantic.com/business/archive/2012/03/take-my-privacy-please-a-defense-of-google/254159/ [https://perma.cc/3B27-DWMB].

what becomes familiar and easily comprehensible to a bench populated by lifelong generalists who chiefly educate themselves through legal briefing and the testimony of paid experts. This reflex produced the broken status quo we are in today.

The need for the fourth party doctrine and inference-facing principle is acute. While regulators and the judiciary play games of catch-up and whack-a-mole, a new generation is maturing with *no* concept of privacy as the Court, in either its majority or dissent, understood it in *Katz*. They are already accustomed to Big Data peering at their search history and predicting their next purchase; they will not notice when equally penetrating government surveillance and investigation grows and pervades the American social fabric. *Carpenter* ruled on a twenty-year-old technology; by the time another *Carpenter* comes along, will we remember what privacy felt like?