Washington Law Review

Volume 95 | Number 2

6-2020

Privacy Dependencies

Solon Barocas
Cornell University, sbarocas@cornell.edu

Karen Levy
Cornell Law School, karen.levy@cornell.edu

Recommended Citation
Solon Barocas & Karen Levy, *Privacy Dependencies*, 95 Wash. L. Rev. 555 (2020).
Available at: https://digitalcommons.law.uw.edu/wlr/vol95/iss2/4

PRIVACY DEPENDENCIES

Solon Barocas[*] & Karen Levy[**]

Abstract: This Article offers a comprehensive survey of privacy dependencies—the many ways that our privacy depends on the decisions and disclosures of other people. What we do and what we say can reveal as much about others as it does about ourselves, even when we don't realize it or when we think we're sharing information about ourselves alone. We identify three bases upon which our privacy can depend: our social ties, our similarities to others, and our differences from others. In a *tie-based dependency,* an observer learns about one person by virtue of her social relationships with others—family, friends, or other associates. In a *similarity-based dependency*, inferences about our unrevealed attributes are drawn from our similarities to others for whom that attribute is known. And in *difference-based dependencies*, revelations about ourselves demonstrate how we are different from others—by showing, for example, how we "break the mold" of normal behavior or establishing how we rank relative to others with respect to some desirable attribute. We elaborate how these dependencies operate, isolating the relevant mechanisms and providing concrete examples of each mechanism in practice, the values they implicate, and the legal and technical interventions that may be brought to bear on them. Our work adds to a growing chorus demonstrating that privacy is neither an individual choice nor an individual value—but it is the first to systematically demonstrate how different types of dependencies can raise very different normative concerns, implicate different areas of law, and create different challenges for regulation.

[*] Assistant Professor, Cornell University, Department of Information Science; Principal Researcher, Microsoft Research; Faculty Associate, Berkman Klein Center for Internet and Society, Harvard University.

[**] Assistant Professor, Cornell University, Department of Information Science; Associated Faculty, Cornell Law School. For helpful feedback and discussions about previous versions of this Article, we thank Joe Calandrino, Ryan Calo, Bobby Kleinberg, Jon Kleinberg, Mark MacCarthy, Kirsten Martin, Arvind Narayanan, Paul Ohm, Natalie Ram, Omer Reingold, Neil Richards, and participants in the 2018 Privacy Law Scholars Conference, the Beyond Differential Privacy workshop at the Simons Institute, and the Carleton College Faculty Seminar. We gratefully acknowledge the John D. and Catherine T. MacArthur Foundation for its support of this research.

I. INTRODUCTION

When two people—let's call them Alice and Bob[1]—interact, and Alice learns something about Bob in the process, Bob may place his faith in Alice that she will not communicate these details to others. Bob's privacy depends, in part, on Alice's behavior: here, her willingness to abstain from speaking about their interactions. While Bob may rely on various social mechanisms—personal requests, social sanctions, harms to Alice's reputation, etc.—to ensure that Alice does not divulge his information to others, Bob cannot exercise complete control over Alice's behaviors. And perhaps he should not be able to; allowing Bob such a right suggests that Alice has no—or perhaps a lesser—claim to those details that emerged in their interaction. When preferences conflict, it can be practically difficult to disentangle whether the information "belongs" to Alice or to Bob and which of them ought to have control over disclosure decisions.[2]

Such conflicts are common on social media, where, for example, Alice may post an unflattering photo of Bob that Bob would rather not have others see.[3] When Alice and Bob disagree on whether the photo should remain online, whose interests should prevail? While social networks like Facebook have carved out important exceptions for when Bob might assert a superior privacy claim over the photo,[4] the platform does not grant Bob exclusive rights over any image in which he might appear—instead, it encourages users like Bob "to get in touch with the person who posted

[1] Alice and Bob are "the world's most famous cryptographic couple." Since 1978, the fictional duo has been used as standard placeholders in explanations of cryptographic protocols and other engineering problems (e.g., "Alice sends a message to Bob," "Alice and Bob wish to exchange a private key"). *See* Quinn DuPont & Alana Cattapan, *Alice and Bob: A History of the World's Most Famous Couple*, CRYPTOCOUPLE (2017), http://cryptocouple.com/Alice%20and%20Bob%20-%20DuPont%20and%20Cattapan%202017.pdf [https://perma.cc/PTZ9-22A4].

[2] *See generally* SANDRA PETRONIO, BOUNDARIES OF PRIVACY: DIALECTICS OF DISCLOSURE (2002) (describing those entrusted with others' private information as "co-owners" of that information).

[3] *See* Gergely Biczók & Pern Hui Chia, *Interdependent Privacy: Let Me Share Your Data*, *in* INTERNATIONAL CONFERENCE ON FINANCE CRYPTOGRAPHY AND DATA SECURITY 338 (2013) (describing nonconsensual photo tagging as an example of privacy's interdependent nature).

[4] These include, among other things, non-consensual pornography. *See Community Standards: Sexual Exploitation of Adults*, FACEBOOK, https://www.facebook.com/communitystandards/safety/sexual_exploitation_adults [https://perma.cc/7NRR-L7N7].

this content in order to resolve the issue."[5] Apple similarly relies on its users to negotiate disparate privacy preferences about whether to submit voicemail recordings to improve its speech recognition algorithms. If Bob leaves a voicemail on Alice's phone, Alice is charged with the decision about whether to give Apple access to it—but is warned: "Do not submit recordings *if you believe the speaker would be uncomfortable* with you submitting the content to Apple."[6] Google likewise invokes interpersonal etiquette, imploring the owners of Nest smart devices to disclose the devices' presence to guests in their homes.[7] By encouraging users to work out privacy conflicts among themselves, rather than mediating the conflict through rule or technology, Facebook, Apple, and Google evince the common hesitation that platforms and policymakers have about involving themselves too directly in what are seen as interpersonal information conflicts.[8]

One might view these situations as a conflict between Bob's privacy and Alice's freedom of speech.[9] For our purposes, they highlight a more fundamental point: to the extent that people do not retreat completely from

[5] *See Photos or Videos that Violate Your Privacy*, FACEBOOK, https://www.facebook.com/help/imageprivacyrights [https://perma.cc/4F3G-U4A2]. Strahilevitz describes these situations as "collective privacy" conflicts. LIOR STRAHILEVITZ, THE OFFENSIVE INTERNET: SPEECH, PRIVACY, AND REPUTATION 217 (Saul Levmore & Martha Nussbaum eds., 2010); James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1171–72 (2009) (describing photo tagging and un-tagging on Facebook due to individuals' divergent privacy preferences). Facebook also throws up its hands with respect to personal information shared through friends' contact lists, stating that: "People own their address books . . . . We understand that in some cases this may mean that another person may not be able to control the contact information someone else uploads about them." Kashmir Hill, *Facebook is Giving Advertisers Access to Your Shadow Contact Information*, GIZMODO (Sep. 26, 2018), https://gizmodo.com/facebook-is-giving-advertisers-access-to-your-shadow-co-1828476051 [https://perma.cc/UJB8-2FWA] (quoting a Facebook spokesperson).

[6] *See* Anthony Bouchard, *How to Use Voicemail Transcription on iPhone,* IDOWNLOAD BLOG (Sep. 21, 2016), https://www.idownloadblog.com/2016/09/21/iphone-voicemail-transcription/ [https://perma.cc/F4CL-3EC3] (emphasis added); Peter Skomoroch (@peteskomoroch), TWITTER (Oct. 10, 2018, 10:30 PM), https://twitter.com/peteskomoroch/status/1050197774430396416 [https://perma.cc/F99M-KYYA] (displaying screenshot of Apple's instructions).

[7] Leo Kelion, *Google Chief: I'd Disclose Smart Speakers Before Guests Enter My Home*, BBC NEWS (Oct. 15, 2019), https://www.bbc.com/news/technology-50048144 [https://perma.cc/6MXV-C38F]. Google's devices chief continued, however, to suggest that data collection is "probably something that the products themselves should try to indicate." *Id.*

[8] *See* Karen Levy, *Relational Big Data*, 66 STAN. L. REV. ONLINE 73, 78 (2013) ("[I]n most cases, interpersonal privacy intrusions . . . fall outside the realm of legal redress, precisely because the law is traditionally hesitant to get involved in the minutiae of personal relationships."); Karen Levy, *Intimate Surveillance*, 51 IDAHO L. REV. 679, 692 (2015) (describing law's hesitation to "rais[e] the curtain upon domestic privacy" by exposing that which "ought to be left to family government" (quoting State v. A.B. Rhodes*,* 61 N.C. (Phil.) 453, 454, 459 (1868))).

[9] Asserting that a person should be able to control what another says about their interactions is what Eugene Volokh has called a "right to stop people from speaking about you." Eugene Volokh, *Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking About You*, 52 STAN. L. REV. 1049, 1049 (2000).

society, *everyone's privacy depends on what others do*. There is no way to live in the world without putting yourself at risk that others might make use of information about you in ways to which you do not consent. This is as true of someone interacting with close family, friends, and colleagues as it is of someone walking down a busy city street among strangers. No one can claim exclusive privilege to the information communicated in these encounters. In this most basic sense, individuals' privacy always depends on others' discretion.

The ever-present possibility that Alice might betray Bob's confidence does not mean that Bob lives in a constant state of anxiety. When the social tie is a close one, social, reputational, and emotional considerations can operate to limit Alice's disclosures about Bob.[10] Information sharing often closely accompanies social connection: pals are commonly confidants,[11] and the fact that Alice and Bob have control over one another's private information can facilitate mutual cooperation, trust, and confidence in their relationship. Bob's dependency on Alice effectively communicates to Alice that he trusts her with such details, fostering intimacy rather than suspicion. Alice may recognize that withholding information shared in confidence is necessary to maintain Bob's confidence in her. And when trading personal information back and forth, Alice and Bob may knowingly and happily put themselves in a position of mutual dependency.[12]

The regulation of Bob's privacy vis-à-vis Alice's behavior, under these conditions, is really a matter of appropriate social conduct, where violations are met with accusations of betrayal or a questioning of character.[13] Social norms can curb information sharing that implicates

[10] Alice E. Marwick & danah boyd, *Networked Privacy: How Teenagers Negotiate Context in Social Media*, 16 NEW MEDIA & SOC'Y 1051, 1061 (2014) ("In a networked setting, teens cannot depend on single-handedly controlling how their information is distributed. What their peers share about them, and what they do with the information they receive cannot be regulated technically, but must be negotiated socially."); Joshua A.T. Fairfield & Christoph Engel, *Privacy as a Public Good*, 65 DUKE L.J. 385, 452 (2015) ("Shaming is a social sanction, which is frequently used as a reaction to informational damage. Spread rumors about a sister-in-law, and expect to be ostracized at family gatherings. Air dirty laundry on Facebook, and expect to be defriended.").

[11] CYNTHIA FEE, THANK YOU FOR BEING A FRIEND (Asylum Records 1985) (theme song for *The Golden Girls*).

[12] ARI E. WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 51 (2018).

[13] As Nissenbaum has argued, privacy should thus not be understood as a matter of control, but rather in terms of information flows that abide by context-dependent norms. HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2009); *see also* Richard H. McAdams, *Group Norms, Gossip, and Blackmail*, 144 U. PA. L. REV. 2237 (1996) (discussing the relationship between law and norms in the context of blackmail); Lior J. Strahilevitz, *Social Norms from Close-Knit Groups to Loose-Knit Groups*, 70 U. CHI. L. REV. 359 (2003); Yu Pu & Jens Grossklags, *Sharing is Caring, Or Callous?*, *in* INTERNATIONAL CONFERENCE ON FINANCE CRYPTOLOGY AND NETWORK SECURITY 670 (2016) (investigating relationships between the value people attach to data about their friends and their perceived level of bonding social capital).

other people's privacy, even in the absence of law—and can do so in ways that are recognized as necessary for a well-functioning society.

But privacy norms, however important, are limited prophylactics to these problems. The situation described above is a simple one: Bob depends on Alice not to tell others his secret. But our privacy is determined by others' choices in many far more complicated situations than this. This may be because Alice's disclosure implicating Bob is involuntary, or because she doesn't know the effect it will have on Bob; or because she has no relationship to or knowledge of Bob at all. In fact, Alice may disclose information that is *explicitly and exclusively about Alice*, seemingly having nothing whatsoever to do with Bob, and can still implicate his privacy in so doing. In these situations, social norms are of limited utility in protecting Bob's privacy.

This Article explores the varied ways in which one person's privacy is implicated by information others reveal. We term these phenomena *privacy dependencies* and we identify three broad types. In a *tie-based dependency*, an observer learns about one person by virtue of that person's social relationships with others—family, friends, or other associates. This may occur, for example, when a person subject to surveillance communicates with others: even those who are not the person of interest might be "caught in the net" of observation. In a *similarity-based dependency*, inferences about our unrevealed attributes are drawn from our similarities to others for whom that attribute is known. And in *difference-based dependencies*, revelations about ourselves demonstrate how we are different from others—by showing how we "break the mold" of normal behavior, showing how we rank relatively on some desirable attribute, or by allowing an observer to pinpoint an unknown person through process of elimination.

Prior research has explored privacy's socially interdependent nature in various ways.[14] Research on the *social value of privacy* underscores the necessity of privacy to social functioning: individual privacy guarantees enable collective values to flourish by making space for individuals to live freely, interact unreservedly, and participate fully in social life. In this way, social groups enjoy the benefits afforded by individual privacy.[15]

[14] For a survey of work about how different technical research communities have examined interdependent privacy and a review of methodologies that have been used to study the problem, see Mathias Humbert et al., *A Survey on Interdependent Privacy*, 52 ACM COMPUTING SURVEYS 122 (2019), https://infoscience.epfl.ch/record/264048 [https://perma.cc/BC8Y-JAAX].

[15] *See generally* PRISCILLA REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY (1995); BEATE ROESSLER & DOROTA MOKROSINSKA, SOCIAL DIMENSIONS OF PRIVACY: INTERDISCIPLINARY PERSPECTIVES (2015); Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421 (1980); Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609 (1999); Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy,* 44 SAN DIEGO L. REV. 745, 760–64 (2007).

Complementary work on *group privacy* recognizes privacy interests that inhere in group membership—both the "right to huddle" in private association with others[16] and a privacy interest in aggregated statistical attributes of groups to which one belongs.[17] Work on *relational privacy* recognizes that different privacy expectations attach to different people and institutions in our lives and suggests that law should take into account these different sensitivities in setting rules about such expectations (for example, by recognizing that we may have a greater interest in privacy against the government than we do against our neighbors).[18] *Networked privacy* explores the complex and creative practices required to negotiate information flows in networked spaces (e.g., social media platforms), thanks in part to others' roles in sharing information about us.[19] Other work has drawn from economic concepts—exploring, for example, the idea of *privacy externalities*, which exist "where one person's decision to share information can adversely affect others who choose to remain silent,"[20] and *privacy as a public good*, which observes that "[a]n

[16] *See* Edward J. Bloustein, *Group Privacy: The Right to Huddle*, 8 RUTGERS-CAMDEN L.J. 219 (1977). Bloustein is often credited with coining the term *group privacy*, which he suggests is "a form of privacy that people seek in their associations with others." *Id.* at 221. On Bloustein's account, group privacy "is an attribute of individuals in association with one another within a group, rather than an attribute of the group itself[,]" akin to the shared secrecy expected between lovers or in a football huddle. *Id.* at 221–23. Bloustein's concern is about individuals associating themselves with others, rather than based on shared characteristics among people.

[17] LINNET TAYLOR, LUCIANO FLORIDI & BART VAN DER SLOOT, GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES (2017); Anton Vedder, *KDD: The Challenge to Individualism*, 1 ETHICS & INFO. TECH. 275 (1999). In light of the limitations of the individual privacy model to address the harms of data mining, Vedder introduces the idea of *categorical privacy*, which protects "information . . . [which is] originally taken from the personal sphere of individuals, and—after aggregation and processing according to statistical methods—is no longer accompanied by identifiers of individual natural persons, but, instead, by identifiers of groups of persons[.]" *Id.* at 279. Taylor et al. discuss two ontologies of group privacy: one attached to groups of individuals (what they term an *entity-first* approach, in which group privacy is understood as "a result of the collection of the privacies of the constituting members") and one attached to particular attributes (a *predicate-first* approach, in which group privacy is "an emergent property, over and above the collection of the privacies of the constituting members"). TAYLOR ET AL., *supra*, at 7–8.

[18] Karen Levy et al., *Regulating Privacy in Public/Private Space: The Case of Nursing Home Monitoring Laws*, 26 ELDER L.J. 323, 327–29 (2019); Laurent Sacharoff, *The Relational Nature of Privacy*, 16 LEWIS & CLARK L. REV. 1249 (2012).

[19] Networked privacy conceptualizes privacy as a set of practices to manage and negotiate boundaries between audiences and contexts in networked information systems. As previously bounded social contexts blur into one another and disrupt context-specific norms, people struggle to control what information is shared about them, and by whom. As a result, people develop new strategies, both individual and collective, to manage their privacy (e.g., steganographic posting, strategic content curation). *See* Eszter Hargittai & Alice Marwick, *"What Can I Really Do?" Explaining the Privacy Paradox with Online Apathy*, 10 INT'L J. COMM. 3737 (2016); Marwick & Boyd, *supra* note 10, at 1603; Phillip Fei Wu et al., *A Contextual Approach to Information Privacy Research*, 70 J. ASS'N INFO. SCI. & TECH. 1 (2019).

[20] Mark MacCarthy, *New Directions in Privacy: Disclosure, Unfairness and Externalities*, 6 I/S: J. L. & POL'Y FOR INFO. SOC'Y 425, 428–29 (2011). Our notion of *privacy dependencies* is somewhat broader than MacCarthy's definition of *privacy externalities*. MacCarthy explicitly excludes from his

individual who is careless with data exposes not only extensive information about herself, but about others as well."[21]

And for perhaps the first time, interdependence is becoming part of mainstream public discourse about privacy. This development owes in large part to recent high-profile situations in which one person's privacy has depended on the choices of another—most notably, the Cambridge Analytica scandal in March 2018 (in which Facebook users unwittingly revealed information *about their friends* to a political consulting firm) and the increasing use of familial DNA search in criminal investigations (in which suspects are apprehended based on DNA their *relatives* submitted to genealogical databases). In the public imagination, these two situations lay bare the degree to which our most intimate associates—friends and family—can expose us.[22] But our privacy depends on others in far more situations than these, and in many diverse forms—including contexts in which informants are more socially distant than friends or family, and even less likely to be governed by relational norms that might mediate disclosure.

Scholars have pointed to privacy's social nature as yet another nail in the coffin of the individualistic, notice-and-consent model of privacy regulation, arguing that in addition to the model's other problems,[23] it fails to provide protection for those whose privacy depends on others but who

analysis "phenomena where one person is directly disclosing information about another person[,]" focusing instead on cases in which "the data subject reveals information only about himself" that nonetheless negatively impacts others. *Id.* at 449. The types of dependencies we describe include instances of both cases. Economists have also studied the issue using game-theoretic approaches, demonstrating that information externalities can lead to market disequilibria. *See* Daron Acemoglu et al., *Too Much Data: Prices and Inefficiencies in Data Markets* (Nat'l Bureau of Econ. Research, Working Paper No. 26296, 2019) (establishing that others' disclosures will undermine consumers' ability to command a price for their own data that is in keeping with how much they value their privacy); Jay Pil Choi et al., *Privacy and Personal Data Collection With Information Externalities*, 173 J. PUB. ECON. 113 (2019) (demonstrating that externalities lead to disequilibrium even with informed consent); Mathias Humbert et al., *On Non-Cooperative Genomic Privacy*, *in* INTERNATIONAL CONFERENCE ON FINANCE CRYPTOGRAPHY AND DATA SECURITY 407 (2015) (showing inefficient equilibria in the context of sharing genetic data when family members have different sharing preferences).

[21] Fairfield & Engel, *supra* note 10, at 385. Our exploration here includes some cases in which an individual is "careless" with his own data and hence implicates others' interests, but also cases in which individuals disclose their own data involuntarily.

[22] As an example of such discourse, see Will Oremus, *How the Golden State Killer's DNA Search is Like the Cambridge Analytica Scandal*, SLATE (May 1, 2018), https://slate.com/technology/2018/05/how-the-golden-state-killers-dna-search-is-like-the-cambridge-analytica-scandal.html [https://perma.cc/M4BB-W9SP] ("Cambridge Analytica and GEDmatch are a stark reminder that the problems go deeper than just better informing users of what they're giving up *about themselves.*" (emphasis added)).

[23] *See generally* NISSENBAUM, *supra* note 13; Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013), for a review of the philosophical and practical grounds for discounting the individual model of privacy regulation.

receive no opportunity to withhold consent.[24] We agree—and claim further that the *particular mechanisms* of the dependency pose meaningfully different threats to privacy and its protection. We survey the various mechanisms behind these dependencies, placing them into a framework that highlights how they relate to and differ from one another. In so doing, we illustrate how different normative values—freedom of association, social solidarity, nondiscrimination, and others—attach to these arrangements, and consider how social practices, policies, and technical interventions respond to them. We pay particular attention to how different dependencies implicate diverse areas of law.

Part II presents this framework. We consider three different categories of privacy dependency: dependencies based on a *tie* between individuals (section II.A); dependencies based on *similarity* between them (section II.B); and dependencies based on *differences* between them (section II.C). In Part III, we explore the interaction of our dependency forms in the context of genetics, as an illustration of how they merge, conflict, and implicate different values in practice. Part IV concludes with implications for privacy's protection.

II. THREE TYPES OF PRIVACY DEPENDENCIES

First, a comment on notation. We'll refer throughout to three characters: Alice, Bob, and the Observer. In our model, Alice is the party who reveals some sort of information to the Observer—and in all cases, Alice's disclosure leads to the Observer learning some information about Bob. We can say, in each case, that Bob's privacy is dependent on Alice's disclosure or nondisclosure of information. We choose not to personalize the Observer beyond its instrumental status as the collector of Alice's information. The Observer could, in principle, be an individual—but as we shall see, in most practical cases, it stands in for a corporate or governmental actor (say, the police or a social media platform).

Across our cases, there is considerable variety about *what* it is that Alice shares and *why* she does so. Alice's disclosure may include information that is—on its face—solely about Alice; solely about Bob; or about the both of them and their relationship to one another. In varying circumstances, Alice may or may not intend, or even know, that she is

[24] *See, e.g.*, MacCarthy, *supra* note 20, at 447 (arguing against the individual informed consent model of privacy protection because "[i]ndividual level choices will result in data collection and use patterns that impose substantial tangible costs on individuals who are not directly involved in making those choices"). MacCarthy argues, instead, for an "unfairness framework" that weighs the public benefit of information practices against prospective harms, and incorporates potential privacy externalities into the calculus. *Id.* at 430. Fairfield and Engel similarly note the limitations of the individual model, and argue for using tools and concepts from the behavioral economics literature on public goods to protect privacy at the group level. Fairfield & Engel, *supra* note 10, at 388–89; *see also* Joel R. Reidenberg et al., *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, 11 I/S: J.L. & POL'Y INFO. SOC'Y 485 (2015).

disclosing anything at all to the Observer—or she may be coerced, incentivized, or required to do so. Alternatively, she may reveal information wholly upon her own volition, with or without awareness of the privacy consequences of her actions for Bob.

As a final wrinkle, a privacy dependency may involve one or many Alices, and one or many Bobs. The Observer may learn something meaningful about Bob only upon the disclosures of several Alices; alternatively, a single Alice may reveal something with privacy consequences for numerous Bobs. And as we shall see, both additional combinations (an Alice and a Bob, many Alices and many Bobs) also occur.

A. Tie

Data about us often reside in those with whom we associate. An Observer who gathers information about Alice may learn about Bob by virtue of his connection to Alice. Indeed, by capitalizing on the relationship between Alice and Bob, an Observer can circumvent obstacles to learning about Bob directly—and can often make better sense of the information obtained about Alice. Observers commonly leverage our interpersonal connections to collect information about individuals via their associates (a dynamic we call *passthrough*); incidentally observe one person in the course of observing another (*bycatch*); identify unknown people based on their relationships with known others (*identification*); and justify the collection of data about the people with whom individuals are connected (*tie-justified observation*).

1. Passthrough

In some cases, Alice may serve as a conduit through which Bob's information is passed to the Observer. As mentioned above, Bob may have previously shared some bit of personal information with Alice, which Alice subsequently passes along. Alice may share Bob's information knowingly, perhaps upon having been coerced or incentivized to do so. The practice of acting as a confidential informant is a classic example: because law enforcement cannot observe Bob directly, it leverages Alice's social tie to Bob to gather intelligence about Bob. The Observer may provide Alice with some sort of favorable treatment in exchange for providing information about Bob—or may exploit existing weaknesses in Alice and Bob's relationship.[25]

[25] *See* Spencer Headworth, *Getting to Know You: Welfare Fraud Investigation and the Appropriation of Social Ties*, 84 AM. SOC. REV. 171, 181 (2019) (noting that people may be motivated to report associates' welfare fraud as "an instrument for personal agendas or a weapon in interpersonal conflicts").

But Alice may also serve as a passthrough for Bob's data in much more commonplace situations, and even without her knowledge. An Observer may, for instance, trick Alice into providing information about one or many Bobs to whom she is connected. A prominent example is social networks' practice of encouraging users to upload contact lists in order to find friends who already use the service—and to solicit participation by those who don't, while building "shadow" profiles of these non-users.[26] If enough of Bob's friends and associates have uploaded their contact lists, the social network will know Bob's precise position in the social graph, despite his steadfast refusal to join the network. Prompting users to share their contact lists is one of the most common "dark patterns"[27] on the web: platforms and messaging apps often mislead users into doing so through sneaky design tactics or promises of a better user experience on the site.[28]

Less intuitively, privacy dependencies can also result when the *Observer* acts as a passthrough; that is, when Bob seeks to share some information with Alice but can only do so by first passing it to the Observer. These situations, increasingly common as personal communications are mediated by platforms, practically *require* Bob to knowingly reveal information about himself to an Observer—because doing so is the only practical way he can communicate with Alice. Bob may find himself under increasing pressure to reveal information about himself as a larger share of his associates begin to communicate through intermediaries with centralized architectures (i.e., when all the connections between nodes pass through a centralized hub). Common

[26] Daniel K. Gillmor, *Facebook is Tracking Me Even Though I'm Not on Facebook*, ACLU: FREE FUTURE BLOG (Apr. 5, 2018, 6:00 PM), https://www.aclu.org/blog/privacy-technology/internet-privacy/facebook-tracking-me-even-though-im-not-facebook [https://perma.cc/VER3-HHBV] ("Facebook uses . . . contact information to learn about people, even if those people don't agree to participate. It also links people together based on who they know, even if the shared contact hasn't agreed to this use. For example, I received an email from Facebook that lists the people who have all invited me to join Facebook: my aunt, an old co-worker, a friend from elementary school, etc. . . . Facebook records this group of people as my contacts, even though I've never agreed to this kind of data collection."); *see also* Iraklis Symeonidis et al., *Collateral Damage of Facebook Third-Party Applications: A Comprehensive Study*, 77 COMPUTERS & SECURITY 179 (2018) (discussing the prevalence of "collateral information collection"—that is, information about a user's friends—across Facebook apps).

[27] *Friend Spam*, DARK PATTERNS, https://darkpatterns.org/types-of-dark-pattern/friend-spam [https://perma.cc/ZGT7-BXHD].

[28] In 2015, LinkedIn settled a lawsuit for $13 million for harvesting contacts from users through such tactics and then spamming those contacts with invitations, seemingly sent by the user, to join the service. John Brownlee, *After Lawsuit Settlement, LinkedIn's Dishonest Design is Now a $13 Million Problem*, FASTCO DESIGN (Oct. 5, 2015), https://www.fastcodesign.com/3051906/after-lawsuit-settlement-linkedins-dishonest-design-is-now-a-13-million-problem [https://perma.cc/VH5W-U37K]. And WhatsApp came under fire from Canadian and Dutch privacy regulators for auto-populating users' contact lists from their phone address books, and for retaining this contact information on their own servers. Chester Wisniewski, *WhatsApp's Privacy Investigated by Joint Canadian-Dutch Probe*, NAKED SECURITY (Jan. 29, 2013), https://nakedsecurity.sophos.com/2013/01/29/whatsapps-privacy-investigated-by-joint-canadian-dutch-probe [https://perma.cc/6R6W-NVSX].

communication infrastructures ranging from telephones to online social networks enjoy so-called "network effects": as a greater number of Bob's associates participate, Bob stands to benefit more from participating as well.[29] Alice and others' decisions to join (and hence share information with) Facebook, for example, may create powerful incentives for Bob to join, especially if Bob would miss out on valued social interactions should he abstain. Their decisions do not compromise Bob's privacy directly; instead, their choices make Bob's refusal to share information more costly. Bob may find that reaching his associates is less convenient, more expensive, or outright impossible if he does not join the platform.[30] In other words, he might feel compelled to participate by the choices that Alice and others have made.

Network effects can make *leaving* a platform difficult as well.[31] Alice and Bob—along with all of their associates—may decide that they no longer want to share their information with Facebook as a condition of communicating with each other, but they cannot replicate the value that Facebook offers unless they all move, collectively, to another platform. Network effects often create a type of collective action problem because no one wants to be the first person to leave the network. At the same time, individual actors may find it difficult to coordinate a wholesale move. Bob may hesitate to leave unless he knows that he'll be able to find Alice elsewhere; he'd likely have even less confidence in his ability to persuade all of his other associates to make the move with him. Once Alice, Bob, and their associates have made the decision to join a network, none of them may be in a position to orchestrate their effective individual or collective departure.

Data portability and interoperability between platforms are often cited as possible solutions to this problem—but their success in practice has

[29] *See generally* DAVID EASLEY & JON KLEINBERG, NETWORKS, CROWDS, AND MARKETS: REASONING ABOUT A HIGHLY CONNECTED WORLD 509–42 (2010). Network effects can be either general (in which only the *number* of additional adopters provides the benefit) or identity-specific (when *who* adopts matters—for example, high-status users, or one's particular group of friends). In other words, network effects may incentivize Bob's participation on a data-gathering platform either because many Alices choose to adopt (general) or because particular Alices of social import to Bob adopt (identity-specific). *See* Paul DiMaggio & Joseph Cohen, *Information Inequality and Network Externalities: A Comparative Study of the Diffusion of Television and the Internet*, *in* THE ECONOMIC SOCIOLOGY OF CAPITALISM 235–36 (Victor Nee & Richard Swedberg eds., 2004).

[30] Various studies have attempted to specify the social and economic costs of opt-out from social networks. *See, e.g.*, GREG NORCIE & L. JEAN CAMP, THE PRICE OF PRIVACY: AN EXAMINATION OF THE ECONOMIC COSTS OF ABSTENTION FROM SOCIAL NETWORKS 2–3 (2015), http://www.ljean.com/files/abstain.pdf [https://perma.cc/AJ5T-QYUV] (explaining how abstention from social network sites may impede access to career opportunities).

[31] Eric P.S. Baumer et al., *Limiting, Leaving, and (re)Lapsing: An Exploration of Facebook Non-Use Practices and Experiences*, *in* PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 3257, 3257 (2013).

been limited.[32] At best, data portability would allow Alice and Bob to find a different Observer to mediate their communication. And even when standards are interoperable, Alice and Bob will have to reveal information to each other's service providers. Consider what happens if Alice uses Gmail: Bob can only reach her by email if he sends his message to Google's email servers (even if Bob is not himself a Gmail user). In other words, Alice's decision to use Gmail requires Bob to disclose information to Google as a condition of communicating with her.[33]

The law has occasionally taken pains to preserve the sanctity of social ties, but in other respects, it has created structures that enable the easy exploitation of ties for the procurement of information. As we have noted, the law is loath to regulate passthrough-based disclosures in many cases, generally preferring that social norms be relied upon to govern disagreements. For certain social ties—between attorney and client, doctor and patient, priest and penitent—the law does take steps to insulate Bob from having his information passed through Alice, in the interest of maintaining the sanctity and confidence of those socially important relations.[34] The law acknowledges that society benefits when people are able to place themselves in positions of extreme vulnerability vis-à-vis one another, and weighs that benefit over access to those confidences for evidentiary purposes.[35] In some cases, the protection afforded to privileged communications is so strong that it cannot be waived *even if* one party wishes to testify against another.[36] Despite this, in other cases,

[32] Arvind Narayanan et al., *A Critical Look at Decentralized Personal Data Architectures*, DATA USAGE MGMT. ON WEB 1 (2012), http://dig.csail.mit.edu/2012/WWW-DUMW/papers/dumw2012_submission_5.pdf [https://perma.cc/J4EK-NRPK].

[33] For example, *Matera v. Google* centered on Google's automated scanning of *non*-Gmail-users' messages when they sent email to a Gmail account. Google scanned the messages in order to target ads to users. Matera v. Google Inc., No. 15-CV-04062-LHK, 2017 WL 1365021 (N.D. Cal. Mar. 15, 2017). Here, Google had access to the nonusers' messages by virtue of their communication with Gmail users. *Id.* at *2. The case eventually settled when Google agreed to stop scanning these messages for advertising purposes; however, Google still scans nonusers' emails for spam and malware detection and to generate Smart Replies. Brian Fung, *Gmail Will No Longer Snoop on Your Emails for Advertising Purposes*, WASH. POST (June 26, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/06/26/gmail-will-no-longer-snoop-on-your-emails-for-advertising-purposes/ [https://perma.cc/56WT-LTNP].

[34] Upjohn Co. v. United States, 449 U.S. 383, 389 (1981); Stein v. Bowman, 38 U.S. 209, 223 (1839) (preventing a wife's testimony against her husband in order to protect "the enjoyment of that confidence which should subsist between those who are connected by the nearest and dearest relations of life. To break down or impair the great principles which protect the sanctities of husband and wife, would be to destroy the best solace of human existence").

[35] Trammel v. United States, 445 U.S. 40, 44 (1980); Hawkins v. United States, 358 U.S. 74, 77 (1958) ("The basic reason the law has refused to pit wife against husband or husband against wife . . . was a belief that such a policy was necessary to foster family peace, not only for the benefit of [the family], but for the benefit of the public as well.").

[36] The privilege for confidential communications in marriage has been so treated. *See, e.g.*, United States v. Neal, 532 F. Supp. 942, 946 (D. Colo. 1982) (barring a wife *who desired to testify* as to

the law sometimes *exploits* passthrough-based disclosures. The third-party doctrine—which holds that individuals have no constitutional expectation of privacy in information they have voluntarily given to another—was originally justified by the premise that being "betrayed by an informer . . . [is] inherent in the conditions of human society. It is the kind of risk we necessarily assume whenever we speak."[37] But many legal scholars have opined that the doctrine makes less sense given the ubiquity of platform-mediated communication, in which we have virtually no choice but to pass information through Observers (who may then, voluntarily or compulsorily, share that information with the government).[38]

2. Bycatch

In commercial fishing, *bycatch* refers to those species caught unintentionally that are not the species being targeted—dolphins in tuna nets, seabirds, undersized fish, and the like.[39] Though fisheries can take steps to try to reduce bycatch, contemporary fishing technology makes some degree of incidental bycatch unavoidable. While in some cases, non-target species can be returned to the ocean unharmed, many others perish in the course of being captured or are subsequently discarded.[40] Data collection bears similarity to commercial fishing: Bob may be "caught in the net" when Alice's data are targeted for collection.

This collection may be foreseeable but incidental: Bob may be a bystander in a photograph taken of Alice, in which the photographer has no real intent to capture Bob's image.[41] This concern has come to the fore recently in debates over the risks and benefits of police body-worn cameras, and whether the public should have access to the footage they capture—which may contain substantial personal information about crime victims, witnesses, and bystanders in addition to that of officers and defendants. Jurisdictions that require body-worn cameras have attempted

marital communications with her husband from doing so, in the interest of preventing the government from "invad[ing] the confidences of marriage to turn those nearest and dearest into informers").

[37] Lopez v. United States, 373 U.S. 427, 465 (1963) (Brennan, J., dissenting).

[38] *See, e.g.*, United States v. Jones, 565 U.S. 400, 417 (2012) (Sotomayor, J., concurring) (noting that the third-party doctrine "is ill suited to 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").

[39] Martin A. Hall, *On Bycatches*, 6 REVS. FISH BIOLOGY & FISHERIES 319, 321 (1996).

[40] *Id.*

[41] *See, e.g.*, Roberto Hoyle et al., *Privacy Behaviors of Lifeloggers Using Wearable Cameras*, *in* PROCEEDINGS OF THE 2014 ACM INT'L JOINT CONFERENCE ON PERVASIVE & UBIQUITOUS COMPUTING 571, 572 (2014) (investigating how bystanders respond to the presence of lifelogging devices and the willingness of lifeloggers to respect the privacy of bystanders).

to strike a balance between releasing data relevant to "target" individuals and protecting the privacy of others in the frame of such video, commonly through redaction and restrictions on public access.[42]

Bycatch can also occur when Alice and Bob share physical space. Perhaps Bob is a short-term guest in Alice's home, which Alice has equipped with smart devices that capture video, audio, or other types of data (say, an Amazon Echo or a Nest Cam).[43] Bob's data may be captured, processed, and transmitted alongside Alice's back to the vendor of the device, very likely without Bob's knowledge or consent—but also, perhaps, without the Observer's specific intent to capture data about Bob.

Of course, sometimes Alice intends for the Observer to see Bob in the footage. Alice might want a home security camera to transmit Bob's doings in Alice's home to the company that provides the equipment—or on to the police. Amazon's Ring video doorbells, for example, transmit recordings back to the company for storage and processing.[44] Notably, Ring includes a companion app called Neighbors that allows users to share information collected from their device with others in the neighborhood, specifically via Amazon,[45] and Amazon may, in turn, pass this information on to the police.[46]

Or perhaps the Observer takes a photo of Alice, knowing that it will capture Bob in the background. The Observer might enlist Alice's help in snapping a less conspicuous picture of Bob by making it appear that the Observer is only interested in Alice. People routinely employ this tactic when trying to photograph celebrities without their express consent or notice—a kind of reverse photobombing.[47]

Given these risks, we might expect the Observer to take special steps to alert Bob to the possibility that his information might be swept up with Alice's. For example, vendors may design devices in such a way that makes Alice's choice to install or use the devices highly conspicuous to

[42] *See* Bryce C. Newell, *Collateral Visibility: A Socio-Legal Study of Police Body Camera Adoption, Privacy, and Public Disclosure in Washington State*, 92 IND. L.J. 1329 (2017); Mary D. Fan, *Privacy, Public Disclosure, Police Body Cameras: Policy Splits*, 68 ALA. L. REV. 395 (2016).

[43] Eric Zeng, Shrirang Mare & Franziska Roesner, *End User Security and Privacy Concerns with Smart Homes*, *in* PROCEEDINGS OF THE THIRTEENTH SYMPOSIUM ON USABLE PRIVACY & SECURITY 65 (2017).

[44] *Ring Smart Doorbell Cameras*, RING, https://shop.ring.com/pages/doorbell-cameras [https://perma.cc/3EJV-5JMN].

[45] *Neighbors by Ring*, RING, https://shop.ring.com/pages/neighbors [https://perma.cc/H9UP-CS6X].

[46] Kate Cox, *It's the User's Fault if a Ring Camera Violates Your Privacy, Amazon Says*, ARS TECHNICA (Nov. 20, 2019), https://arstechnica.com/tech-policy/2019/11/cops-can-keep-ring-footage-forever-share-it-with-anyone-amazon-confirms [https://perma.cc/N6V8-N9FS].

[47] Gordon Fletcher & Anita Greenhill, *Photobombing: Mobility, Humour and Culture*, *in* PROCEEDINGS OF THE CONFERENCE ON CULTURAL ATTITUDES TOWARD COMMUNICATION AND TECHNOLOGY 198–206 (2010).

Bob.[48] A bright red light may draw Bob's attention to the device, alerting him to the fact that he is being recorded. In other cases, the burden might be placed on Alice to make sure that the implications of her choices are apparent to Bob. Ring, for example, "includes a door/window sticker in the box with each device that is equipped with audiovisual recording capabilities," enjoining—but not requiring—customers to use these visuals to alert people who might be captured by the device.[49] Likewise, the decision to install a smart device in a multi-occupant home may rest with one resident—Alice—even when it implicates others' privacy. In practice, Bob rarely has any say in Alice's decision.[50]

In some special cases, the law has viewed each occupant as having an important claim to preserving the privacy of a shared space. In nursing homes, for example, family members often want to be able to monitor their loved ones remotely, via webcam—even when residents share rooms. State statutes that address nursing home monitoring require that a resident's roommates consent to monitoring, often allowing them to place limits on its use (e.g., restricting the times of day a camera is on) or to switch rooms if they do not consent.[51] In this case, Alice's relatives' interests are not allowed to trump Bob's privacy concern.[52]

Others have proposed alternative legal approaches to protect "secondary users."[53] For example, companies could be required to distinguish between the data of primary users (who ostensibly consented to data collection) and secondary users whose data were collected incidentally. Secondary user data might be subject to shorter data retention periods; the sale of data to third parties that includes secondary

[48] *See* Ryan Calo, *Against Notice Skepticism in Privacy (And Elsewhere)*, 87 NOTRE DAME L. REV. 1027 (2012) (proposing means of providing visceral notice of privacy-invasive technologies).

[49] Letter from Amazon to Sen. Edward Markey (Nov. 1, 2019), https://www.markey.senate.gov/imo/media/doc/Response%20Letter_Ring_Senator%20Markey%2011.01.2019.pdf [https://perma.cc/C2JY-VQE7].

[50] *See* Christine Geeng & Franziska Roesner, *Who's In Control? Interactions in Multi-User Smart Homes*, *in* PROCEEDINGS OF THE 2019 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 268 (2019) ("[R]oommates resolve conflicts by deferring to the default control and agency of the device's installer or owner, a recurring theme in our findings.").

[51] Lipton, *supra* note 53, at 422–24.

[52] Levy et al., *supra* note 18, at 352–55. In practice, however, these laws may not be effective means of protecting roommates' privacy, as they often depend on frequent adjustments that may not be realistic to enact in understaffed nursing homes. *Id.* at 354–55.

[53] Alex B. Lipton, *Privacy Protections for Secondary Users of Communication-Capturing Technologies*, 91 N.Y.U. L. REV. 396 (2016) (discussing inefficacy of current privacy regimes for protecting the interests of secondary users). *See generally* Mariella Dimiccoli et al., *Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation*, 1 PROC. ACM ON INTERACTIVE, MOBILE, WEARABLE & UBIQUITOUS TECHS. (2018) (proposing technical means of blurring bystanders' faces in photos). *See also* Meg L. Jones, *Privacy Without Screens & the Internet of Other People's Things*, 51 IDAHO L. REV. 639 (2014).

users might be prohibited; and secondary users' data might be anonymized using various technical methods.

But for the most part, the law does not vest Bob with privacy protections when he is caught up in the net of Alice's data. Fourth Amendment standing doctrine maintains that a defendant may only challenge a search when one's *own* person or property are searched, under the justification that one has no reasonable expectation of privacy in the person or property of another person.[54] Bob has no recourse, then, against a warrantless search of Alice's car that happens to reveal incriminating evidence about Bob—regardless of whether Alice's own privacy rights were violated in the course of the search.[55] The same logic has been applied to "incidental overhear" of electronic communications with targets of lawful surveillance. In *United States v. Hasbajrami*,[56] the Second Circuit held that a U.S. resident had no Fourth Amendment interest in his emails with a foreign person targeted under section 702 of the Foreign Intelligence Surveillance Act, despite the fact that such collection would foreseeably capture the communications of people other than the target.[57]

Ironically, respecting Bob's privacy when he appears alongside Alice may require that Alice make herself uniquely identifiable: to ensure that no one but Alice's data are captured, an Observer may use biometric tools (like face or voice recognition) to distinguish Alice from anyone else. But doing so creates a "privacy-privacy tradeoff" in which Alice must render *more* information about herself for Bob's privacy to be protected.[58]

3. *Identification*

We can also *identify* an unknown Bob by virtue of his connection to a known Alice. In familial search procedures, unidentified DNA evidence

[54] *See* Rakas v. Illinois, 439 U.S. 128 (1978).

[55] *Id.*

[56] 945 F.3d 641 (2d Cir. 2019).

[57] *Id.* at 662–64. In the foreign surveillance context, data minimization requirements are intended to "to some degree compensate for the possibility of broad incidental collection." PRIVACY & CIVIL LIBERTIES OVERSIGHT BD., REPORT ON THE SURVEILLANCE PROGRAM OPERATED PURSUANT TO SECTION 702 OF THE FOREIGN INTELLIGENCE SURVEILLANCE ACT 116 (July 2, 2014), https://www.pclob.gov/library/702-Report.pdf [https://perma.cc/SC6N-LLBY]. Indeed, in *Hasbajrami*, while the court found that Hasbajrami had no Fourth Amendment interest in the incidental *collection* of his emails, it left open the question of whether *querying a database* of incidentally collected communications violated Hasbajrami's privacy interests. *Id.* at 646.

[58] David E. Pozen, *Privacy-Privacy Tradeoffs*, 83 U. CHI. L. REV. 221, 229 (2016) (discussing cases in which privacy in one respect is traded off against privacy in another—including cases in which "privacy burdens or benefits [shift] from one group . . . to another"). Privacy-privacy tradeoffs are their own kind of dependency, in which Alice must be willing to sacrifice her own privacy to maintain Bob's. In addition to trading off against Alice's privacy, incentivizing companies to develop better biometric recognition tools may impose additional net costs on consumers and their privacy. *See* Lipton, *supra* note 53, at 423.

from a crime scene is compared to identified DNA samples in order to assess whether a person genetically related to (i.e., a family member of) someone of known identity is likely to have committed the crime in question.[59] Here, genetic ties are the basis for dependency: Bob's identification by law enforcement rests on Alice's (voluntary or involuntary) provision of her own DNA. In some cases, Alice's DNA is collected strategically in order to confirm suspicion of a particular Bob. The "BTK" serial murderer, Dennis Rader, was identified based on a match between crime scene evidence and his daughter's DNA, collected without her knowledge by police from a Pap smear she had at a state university hospital five years earlier;[60] police already suspected Rader of the murders based on other evidence, and his daughter's DNA sample confirmed their suspicions.[61] And an arrest recently made in the long-cold case of the Golden State Killer, who raped and murdered numerous victims in California between 1976 and 1986, was based in large part on comparing unidentified crime-scene DNA to a sample submitted to a genealogy website.[62]

Increasingly, DNA samples from crime scenes are run "blind" against large databases of samples from those convicted or arrested without

[59] Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L.R. 291 (2010); *see also* Danah Boyd & Karen Levy, Networked Rights and Networked Harms (working paper, Mar. 14, 2016) (on file with author).

[60] Ellen Nakashima, *From DNA of Family, A Tool to Make Arrests*, WASH. POST (Apr. 21, 2008), http://www.washingtonpost.com/wp-dyn/content/article/2008/04/20/AR2008042002388.html [https://perma.cc/6SKC-VLNB].

[61] Mark Hansen, *How the Cops Caught BTK*, A.B.A. J. (May 1, 2006), http://www.abajournal.com/magazine/article/how_the_cops_caught_btk/ [https://perma.cc/6TU3-FRUL].

[62] Sam Stanton, *Relative's DNA from Genealogy Websites Cracked East Area Rapist Case, DA's Office Says*, SACRAMENTO BEE (Apr. 26, 2018), http://www.sacbee.com/latest-news/article209913514.html (last visited Apr. 13, 2020). Before the arrest, Michelle McNamara, who investigated the Golden State Killer's case for years, lamented the fact that genealogy websites had to date failed to cooperate with law enforcement efforts. Jeva Lange, *Michelle McNamara's Tantalizing Roadmap for Finding a Long Lost Serial Killer*, WEEK (Mar. 19, 2018), http://theweek.com/articles/761206/michelle-mcnamaras-tantalizing-roadmap-finding-long-lost-serial-killer [https://perma.cc/GA2Z-MDNL] ("The most frustrating detail of all is that police have the Golden State Killer's DNA, but they can only compare it to DNA in their database of criminals convicted of felonies—where there are, naturally, no hits. There are other DNA databases in existence, though: 23andMe has 1.5 million people's profiles, and Ancestry.com has 2.5 million. 'Unfortunately,' write McNamara's editors, 'neither company will work with law enforcement, citing privacy issues and their terms of service.' Yet '[i]f we could just submit the killer's actual genetic material . . . to one of these databases, the odds are great that we would find a second or third cousin and that person would lead investigators to the killer's identity.'"). Those databases have since grown to about 5 million and 10 million profiles, respectively. Gina Kolata & Heather Murphy, *The Golden State Killer is Tracked Through a Thicket of DNA, and Experts Shudder*, N.Y. TIMES (Apr. 27, 2018), https://www.nytimes.com/2018/04/27/health/dna-privacy-golden-state-killer-genealogy.html [https://perma.cc/E9DG-FR8Y].

conviction,[63] without particularized suspicion as to the suspect's identity. Here, establishing a relationship between a crime-scene sample and a known Alice in the database serves as an investigative tool to locate likely perpetrators.[64] In each case, the genetic link between Alice and Bob allows an Observer (here, the state) to identify Bob *through* Bob's connection to Alice—because Bob is unknown to the Observer, or because direct access to Bob is legally or practically unavailable.

Traditional police work also exploits this possibility all the time, initiating the search for an unknown suspect by investigating the suspect's known associates. Consider the following situation: Alice and Bob appear in surveillance footage together. The police recognize Alice by sight, given their prior interactions with her, but they do not recognize Bob, given his lack of criminal history. Yet simply observing Alice and Bob together gives the police a clue about Bob's identity because the police will exploit their knowledge of Alice to identify Bob—her apparent associate.

Social networks follow a similar strategy when they try to identify the people that appear in uploaded photos using facial recognition technology. Rather than attempting to match someone's face to the face of all users on the social network (which may include many millions of people), service providers will frequently limit the set of possible candidates to the known associates of the person who has uploaded the photo (which is more likely to be in the hundreds). Limiting the candidate pool makes the task of comparing faces much less challenging, computationally speaking, and

[63] Solomon Moore, *F.B.I. and States Vastly Expand DNA Databases*, N.Y. TIMES (Apr. 18, 2009), https://www.nytimes.com/2009/04/19/us/19DNA.html [https://perma.cc/5NGE-CLRH]. Forensic DNA databases significantly overrepresent Black and Latino individuals due to disparities in arrest and imprisonment, leading to concerns about further racial profiling and deepening inequities in criminal justice contact rates. *See* Peter A. Chow-White & Troy Duster, *Do Health and Forensic DNA Databases Increase Racial Disparities?*, 8 PLOS MED. (2011).

[64] Recent scholarship has explored the constitutionality of such procedures. *See* David Kaye, *The Genealogy Detectives: A Constitutional Analysis of "Familial Searching"*, 51 AM. CRIM. L. REV. 109 (2013); Murphy, *supra* note 59. Practicing lawyers have begun to raise such objections, as in the 2019 case of Jesse Bjerke, identified through familial search as the prime suspect in a 2016 rape in Alexandria, Virginia. Bjerke's defense attorney sought to have the DNA evidence that lead to his identification excluded, arguing that the act of extracting genetic information from Bjerke's discarded materials constituted a search that should have required a warrant. Note that his attorney did not raise a constitutional challenge to the search against the genetic database (i.e., against others' data); the focus was on the collection of his own data. To date, lawyers have not tested the constitutionality of the matching process itself. Rachel Weiner, *Alexandria Rape Suspect Challenging DNA Search Used to Crack Case*, WASH. POST (June 10, 2019), https://www.washingtonpost.com/local/public-safety/alexandria-rape-suspect-challenging-dna-search-used-to-crack-case/2019/06/10/24bd0e34-87a5-11e9-a870-b9c411dc4312_story.html [https://perma.cc/D94T-7RXH]. For now, the police's ability to perform familial searches is only limited by the terms of service set by the operators of genetic databases, which has been a major point of controversy. States like Maryland have begun to consider legislation that addresses this problem head-on. Megan Molteni, *Should Cops Use Family Tree Forensics? Maryland Isn't So Sure*, WIRED (Feb. 6, 2019), https://www.wired.com/story/maryland-considers-banning-genetic-genealogy-forensics/ [https://perma.cc/WV8R-NJUG].

cuts down on the risk of false matches, which grows as more candidates are considered.[65] To return to our familiar characters, the social network might exploit the fact that Bob belongs to Alice's articulated social network—that they are "friends"—to increase the likelihood of correctly matching a face in one of Alice's uploaded photos to Bob. Because users are more likely to take photographs of and with their friends than with strangers, facial recognition is markedly improved when it integrates information about a user's social network.[66] Knowing that Alice and Bob are friends makes it far easier for a social network to identify Bob by sight when he appears in Alice's photos.

4. Tie-Justified Observation

Social ties between Alice and Bob may be used to justify expanding the scope of surveillance from an initial focus on Alice to also include Bob. Here, the Observer's aim is often to build a broadly inclusive database rather than to learn about any specific user; in other words, the value of observing Alice stems from Alice's connection to *many* Bobs, rather than to a specific Bob. For example, the National Security Agency's bulk telephony metadata program relied on a practice known as "contact-chaining," analyzing phone records that were up to three degrees of separation—or "hops"—out from a suspected terrorist.[67] Effectively, this allowed the NSA to collect data from about 20 million people for each initial target.[68] (In 2014, President Obama limited the scope of inquiry to two degrees of separation away from a suspect seed—approximately 25,000 people per suspect.)[69] The three-hop (and then two-hop) rule relied explicitly on a network tie between Alice and Bob as a justification for data collection about Bob.

[65] Lucas Introna & Helen Nissenbaum, *Facial Recognition Technology: A Survey of Policy and Implementation Issues*, CTR. CATASTROPHE PREPAREDNESS & RESPONSE 3 (2009), https://nissenbaum.tech.cornell.edu/papers/facial_recognition_report.pdf [https://perma.cc/9FSB-PJ9D] ("Given that the number of possible images that enter the gallery as near-identical mathematical representations (biometric doubles) increases as the size of the gallery increases, restricting the size of the gallery . . . may help maintain the integrity of the system and increase overall performance.").

[66] ZAK STONE ET AL., AUTOTAGGING FACEBOOK: SOCIAL NETWORK CONTEXT IMPROVES PHOTO ANNOTATION, IEEE COMPUTER VISION & PATTERN RECOGNITION WORKSHOPS (2008).

[67] David S. Kris, *On the Bulk Collection of Tangible Things*, 7 J. NAT'L SEC'Y L. & POL'Y 209 (2014).

[68] Amy Nordrum, *NSA Can Legally Access Metadata of 25,000 Callers Based on a Single Suspect's Phone*, IEEE SPECTRUM (May 16, 2016), https://spectrum.ieee.org/tech-talk/telecom/security/nsa-can-legally-access-metadata-of-25000-callers-based-on-a-single-suspects-phone-analysis-suggests [https://perma.cc/H7AB-SN4H].

[69] *Id.* Even under the two-hop rule, targeting only 1% of individuals allows the observer access to 46% of all communications in the network. Laura Radaelli et al., *Quantifying Surveillance in the Networked Age: Node-based Intrusions and Group Privacy*, ARXIV (2018), https://arxiv.org/pdf/1803.09007.pdf [https://perma.cc/J33D-GLSF].

Online social networks often use social ties to set conditions of visibility for their users. For example, Facebook's former privacy settings were premised on network structure: a previous default setting allowed friends of your friends (i.e., people two degrees from a given person) to see certain components of that person's profile.[70] The company employed the same reasoning in determining which information would be available to other actors on the platform: at one point, third-party apps were allowed to collect data from a consenting user's (presumably non-consenting) *friends*. In the Cambridge Analytica scandal, 270,000 Facebook users took a personality quiz through a Facebook app; the app then collected personal data not only from those directly observed users, but also from those users' friends—widening the net to an estimated 87 million indirectly observed users.[71]

Tie-justified observation allows the Observer to learn about many Bobs because they happen to be connected to Alice. Two- or three-hop searches justified by such connections can give Observers permission to learn about far more Bobs than one might expect if the social network includes a number of high-degree nodes—that is, a person in the network that has a particularly large number of connections.[72] If Alice only has a small number of friends, but Bob has a huge number of friends, Alice's connection to Bob would allow an Observer to justify searching Bob's far larger network as well. And if a large proportion of a social network is less than two or three hops from someone like Bob, an Observer might find that they are able to investigate most of the network without having to make any further hops. Effectively, tie-justified observation can easily create a dragnet, in which the number of observed parties far exceeds the number of initial targets.

In the most sympathetic reading, an ostensible justification for such collection is that researchers interested in issues involving social networks—information contagion, the influence of network position on behavior, etc.—can learn much about Alice from learning about those to whom she is connected; in this sense, collection of information about the Bobs to whom Alice is connected is a way of deepening the Observer's knowledge of Alice. In the NSA contact-chaining case, for example, expanding the scope of surveillance from merely a target to a target's

[70] Grimmelmann, *supra* note 5, at 1158.

[71] David C. Vladeck, *Facebook, Cambridge Analytica, and the Regulator's Dilemma: Clueless or Venal?*, HARV. L. REV. BLOG (Apr. 4, 2018), https://blog.harvardlawreview.org/facebook-cambridge-analytica-and-the-regulators-dilemma-clueless-or-venal/ [https://perma.cc/3GH2-VUZ6].

[72] The existence of high-degree hubs (like voicemail services with millions of users) facilitates the scale of observation premised on network tie by creating a large number of new routes to observation via a single phone number. Jonathan Mayer, *MetaPhone: The NSA Three-Hop*, WEB POL'Y BLOG (Dec. 9, 2013), http://webpolicy.org/2013/12/09/metaphone-the-nsa-three-hop/ [https://perma.cc/N4CA-Q526].

associates (and those associates' associates) was controversially justified by virtue of the *relevance* of those parties' communications to the target.[73] But this expansive notion of relevance was used to justify even wider observation: that associates of a known suspect are likely to have additional associates of their own who are worthy of suspicion. Collecting such information was seen as necessary to uncover a coordinating group of suspects (for example, a terrorist cell).[74] However, in its review of the program, the Privacy and Civil Liberties Oversight Board argued that too many "hops" reduces the relevance of the average tie to the initial target, effectively amounting to dragnet collection.[75]

* * *

Observation premised on tie varies in terms of the *nature of the association* between Alice and Bob (genetic; articulated; based on physical proximity or communication), the *intentionality* with which Bob's data are observed via Alice (from incidentally being "caught in the net" of observing Alice, on one end of the spectrum, to purposive and coercive circumvention of obstacles to observing Bob directly, on the other), and the *specificity* with which Bob is targeted by the Observer (whose goal may be to obtain information about a particular Bob or to build a broad social graph of many Bobs). But in all cases, the Observer learns about Bob by "piggybacking" on the tie between Alice and Bob.

Tie-based dependencies tend to implicate the perceived sanctity of our relations with family and friends; our connections to other people, whether severable or persistent, ought not be the source of unconstrained and unanticipated privacy violation. The notion that law enforcement

[73] Robert Chesney, *Telephony Metadata: Is the Contact-Chaining Program Unsalvageable?*, LAWFARE BLOG (Mar. 6, 2019), https://www.lawfareblog.com/telephony-metadata-contact-chaining-program-unsalvageable [https://perma.cc/8NMK-9JSV] (explaining expansion of bulk collection program from "records [pertaining] directly to the agent of a foreign power as defined in FISA, as opposed to spouses, friends or others whose records might well be relevant too[]" to authority where "materials did not belong to or pertain to the particular target (and thus the provision might be used to gather records about *an associate of* the target)" (emphasis added)). *Id.* Chesney goes on to note the faultiness of the relevance justification, which the government argued justified collection of an *entire* comprehensive database as necessary to enable contact-chaining. *Id.*

[74] ADMINISTRATION WHITE PAPER, BULK COLLECTION OF TELEPHONY METADATA UNDER SECTION 215 OF THE USA PATRIOT ACT 4 (Aug. 9, 2013), https://fas.org/irp/nsa/bulk-215.pdf [https://perma.cc/QC9B-LJGL] ("Following the trail [via contact-chaining] . . . allows focused inquiries on numbers of interest, thus potentially revealing a contact at the second or third 'hop' from the seed telephone number that connects to a different terrorist-associated telephone number already known to the analyst.").

[75] PRIVACY & CIVIL LIBERTIES OVERSIGHT BD., REPORT ON THE TELEPHONE RECORDS PROGRAM CONDUCTED UNDER SECTION 215 OF THE USA PATRIOT ACT AND ON THE OPERATIONS OF THE FISC 171 (2014), https://www.pclob.gov/library/215 Report_on_the_Telephone_Records_Program.pdf [https://perma.cc/Y6DE-RPVB] ("Each additional hop from the original 'selector' makes the connection more remote and adds exponentially greater numbers of 'false positives' to the query results. The value of connections becomes more limited as the contact chain is extended and it becomes more difficult to sift through the results.").

might exploit our families' genetic linkages to us, or that a social network might seize upon our articulated friendships to glean our data, may strike us as a profane intrusion into sacred territory. When surveillance is premised on whom you know, we are concerned about the degree to which it impinges on your right to associate freely.[76]

As explained, some information leakage through our relationships is an inevitable side effect of social life—but social norms and interpersonal sanctions have long been relied upon as adequate regulatory mechanisms to restrict unwelcome exposure. Close social ties can both limit and facilitate disclosure, and relationships may operate under the "shadow" influence of a potential Observer. But these considerations do not apply when Alice's disclosure about Bob is unknowing or involuntary—Dennis Rader's daughter almost surely did not consider that her Pap smear would eventually provide incriminating evidence against her father;[77] Cambridge Analytica's quiz-takers are unlikely to have considered the privacy interests of their friends when they divulged their data.[78] When observation depends on the existence of a connection between Alice and Bob rather than on disclosure of the particular information they share (e.g., the NSA's collection of telephony metadata rather than content), the notion that social norms will limit disclosure is similarly weak.[79]

B. *Similarity*

Inference provides yet another—and more circuitous—route to learn about Bob through Alice. When Alice discloses information about herself, she may reveal certain things about Bob as well, if Bob is understood to be *similar* to Alice. That is, the Observer might conclude that Bob likely shares the trait disclosed by Alice if the Observer already knows that Alice

[76] Katherine J. Strandburg, *Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance*, 49 B.C. L. REV. 741, 745 (2008).

[77] Natalie Ram suggests looking to the law of tenancy by the entirety for a framework on how to reconcile multiple (and potentially divergent) interests in shared genetic information. Like genetic relations, tenancy by the entirety is a form of property right in which ownership inheres neither in one spouse nor the other, but in the couple together. Ram draws from this arrangement by analogy in proposing a framework for genetic data. *See* Natalie Ram, *DNA by the Entirety*, 115 COLUM. L. REV. 873 (2015).

[78] David Vladeck's argument that friends should be seen as users too: "The FTC's investigation will focus on a user's reasonable expectations – that is, what did users *and their friends* understand their 'privacy' settings to mean? Were users clearly and unmistakably informed that permitting sharing with friends meant broad and virtually unrestricted access to their data by third parties? Did the 'friends' understand the breadth of third-party access to their data based on decisions that others made?" Vladeck, *supra* note 71.

[79] Despite this, Facebook makes an implausible appeal to norms when urging users to upload their contact lists, informing users on the "Learn More" screen that "You may have business or personal contacts in your phone . . . . Please only send friend requests to people you know personally who would welcome the invite." Kashmir Hill, *How Facebook Figures Out Everyone You've Ever Met*, GIZMODO (Nov. 7, 2017), https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691 [https://perma.cc/L9FD-MVK8].

and Bob resemble each other in many other respects. In this case, the Observer has drawn an inductive inference from its encounter with Alice—that people with a certain set of observable characteristics will also have the disclosed trait. When confronted by Bob, the Observer might recognize that Bob shares these observable characteristics and that Bob is therefore likely to have this additional, unobserved trait as well. Note that the Observer needs to know *something* about Bob in order to apply lessons that it has drawn from Alice and others—that is, whether Bob possesses those characteristics that have tended to *correlate* with the trait in question among other people the Observer has encountered in the past. Unlike information about the sought-after trait, these characteristics might be much more readily observed or more freely disclosed. In fact, these characteristics might be difficult or impossible to conceal or they might seem much more innocuous and thus less worthy of privacy protection.[80] This allows the Observer to sidestep the task—and associated difficulties and discomforts—of observing or asking about the trait directly; instead, the Observer can make a statistically-motivated guess that Bob is likely to follow the same pattern as other people with his same observable characteristics.[81] The ability to draw and apply such inferences means that Alice's disclosures can implicate those similar to her. And it means that observations of Alice can be brought to bear on others who happen to share other known characteristics.

We might understand this dynamic as a form of generalizing, profiling, or stereotyping, where the expectations that the Observer might have about Bob depend on general lessons drawn from particular examples involving many similar Alices. After interacting with many lawyers, for example, an Observer might draw the general conclusion that all lawyers are dishonest. Upon learning that Bob is a lawyer, the Observer might doubt his honesty, too, even though the Observer has no personal experience with him lying. While this conclusion might feel unfounded or objectionable, inductive reasoning of this sort is foundational to human cognition: observations about a specific individual are *always* made meaningful against a backdrop of experience with others.[82] Learning that someone is a lawyer is of no significance to the Observer unless the

[80] Solon Barocas, *Panic Inducing: Data Mining, Fairness, and Privacy* 73–74 (2014) (unpublished Ph.D. dissertation, New York University) (on file with authors).

[81] *See* Solon Barocas, *Leaps and Bounds: Toward a Normative Theory of Inferential Privacy* 17–21 (2015) (unpublished manuscript) (on file with authors).

[82] FREDERICK F. SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 65–67 (2009).

Observer can channel its prior history with lawyers to give substance to the category.[83]

Years ago, Tal Zarsky pointed out that a customer who reveals their shopping patterns on a website may implicate a non-customer who shares the customer's gender, wealth, and zip code because the website can use a model to infer that the non-customer will have similar shopping preferences.[84] Any individual's privacy ultimately depends on what similar people are willing to disclose or what has otherwise been learned about them.[85] In prior work, we showed that this dynamic can lead to a tyranny of the minority, whereby a small number of willing disclosers might determine what observers can then infer about the broader populations to which they belong.[86] Even if a majority of people abstain from disclosing such details or go out of their way to evade observation, the minority of people who happily give up some information may allow observers to uncover the more easily observable or readily disclosed facts that serve as reliable proxies for the sought-after details.[87]

In many respects, this is the goal of a well-executed scientific study: drawing generally applicable conclusions from a limited sample.[88] When

[83] Of course, the stereotype could also have little or nothing to do with the Observer's experience with Alice and those like her. Instead, the stereotype might reflect the Observer's prejudices and biases, loosely connected to or completely detached from any personal encounters with Alice and others like her—gross generalizations or unfounded conclusions. Such stereotypes are rarely learned from direct contact with people so stereotyped; rather, stereotypes tend to gain purchase and their broader cultural force through everyday communication and media representations. Even so, the Observer's ability to assign significance to certain attributes held by Bob will depend on the assertions and representations that others have made about people like Alice.

[84] Tal Z. Zarsky, *Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society*, 56 ME. L. REV. 13, 43–44 (2004) (describing how one customer who reveals his shopping patterns on a website may implicate a non-customer who shares the customer's gender, wealth, and zip code, because the website can use a model to infer that the non-customer will have similar shopping preferences). Mark MacCarthy has likewise explained that "[d]ata from people who have revealed information about themselves through surveys, transactions, and other voluntary disclosures are part of the evidentiary basis for the knowledge revealed by [data mining] techniques. But they apply to other people who have never disclosed that information about themselves." MacCarthy, *supra* note 20, at 425.

[85] Fairfield & Engel, *supra* note 10, at 405 ("Aggregated data contributions serve to train machine learning algorithms, such that the data offered by one person trains an algorithm that impacts someone else.").

[86] Solon Barocas & Helen Nissenbaum, *Big Data's End Run around Anonymity and Consent*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD 44, 61–63 (Julia Lane et al. eds., 2014).

[87] *Id.*

[88] In fact, principles from research ethics demand that researchers enlist the minimal number of research subjects necessary to produce reliable and generalizable findings. Bruno M. Cesana & Paolo Antonelli, *Sample Size Calculations in Clinical Research Should Also Be Based on Ethical Principles*, 17 TRIALS 149, 149 (2016) ("Most statistical methods implemented in controlled clinical trials (CCTs) have the aim of reducing the number of enrolled patients. This aim not only meets the a priori imperative of exposing the minimum number of patients to the burdens of a trial but also fulfills the a posteriori imperative that as few patients as possible are administered the treatment that proves to be inferior." (footnote omitted)); *see also* Peter Bacchetti et al., *Ethics and Sample Size*, 161 AM. J. EPIDEMIOLOGY 105, 106 (2005).

the voluntary disclosures of a willing group of research subjects benefit a broader population, these benefits could be viewed as positive externalities. When these discoveries place others' privacy at risk, the same results could be considered negative externalities. In the first case, Bob might benefit from Alice's decision to participate in a medical study that results in findings applicable to Bob because he shares relevant characteristics with Alice. In the second case, Bob may lose the ability to withhold certain information about himself because Alice has divulged certain medical details and because Bob shares the readily observable characteristics discovered to correlate with Alice's disclosed condition. Of course, the same discovery could play both roles at the same time—helping Bob in some cases and harming him in others.[89]

In most cases, Bob will have no way of keeping Alice from disclosing details that implicate him. Nor will he have any way to prevent the Observer from using Alice's disclosures to draw inferences about him.[90] And yet differences in the dimensions through which the Observer views Alice and Bob as similar can have profound implications for the perceived legitimacy of these inferences. Ultimately, generalizations, profiles, and stereotypes all rest on identifying what is *relevantly* similar about the people who exhibit some quality. The basis upon which a person is understood to be similar can feel extremely proximate or quite distant. At their closest, inferences about your character might depend upon *the company you keep*. Or inferences might depend on the *socially salient characteristics* that you share with others (e.g., gender, race, and age), but with whom you hold no explicit social ties. More distantly, inferences might rest on *characteristics that have little social salience* (e.g., your

[89] This differs from traditional concerns about "group privacy," which have focused on the risks of reporting aggregate statistics from research studies, especially those involving stigmatized medical conditions. If Alice and Bob are known to have participated in a study, then even general statements about the research population can reveal something about them. For example, reporting that 40% of the population suffers from depression means that an Observer can now make a reasoned guess that Alice and Bob have a 40% chance of being depressed. If researchers further report the different rates at which men and women in the study suffer from depression, then the Observer can use Alice and Bob's gender to make an even more precise guess about their probability of being depressed. Now consider what happens if researchers present these as *generalizable* findings rather than aggregate statistics about a particular set of research subjects. Suddenly, *all* men and women might be subject to this inference as well, even if they were not involved in the study. *See* MacCarthy, *supra* note 20, at 458 (describing concerns that medical testing on individual Ashkenazi Jews, who gave voluntary consent, could negatively implicate anyone from such group, because the results could be applied to stigmatize Ashkenazi Jews more generally).

[90] *See, e.g.*, Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 543 (2019) (explaining how the GDPR fails to provide meaningful protection against inference).

preferred web browser[91]) but nevertheless distinguish the group to which you belong from others with respect to the sought-after quality. In each case, making sense of any particular observation or disclosure ultimately depends on inferences rendered on the basis of knowledge about others. And yet, how we are seen as relevantly similar will often dictate whether we are willing to accept or tolerate the resulting inference.

1. The Company You Keep

Consider the inferential bridge that social relationships might provide an Observer. We might infer, for example, that members of the same family are likely to resemble one another along some dimension ("the apple doesn't fall far from the tree"). Given the Observer's knowledge about Alice and the discovery that Bob is Alice's son, the Observer might conclude that Bob is likely to possess many of the same qualities that Alice is known to possess. Such reasoning might rest on the knowledge that certain traits are hereditary, but it could also depend on the belief that parents impart certain qualities to their children through their upbringing. The COMPAS tool—the subject of a well-known ProPublica investigation of "machine bias"[92]—asks defendants about the criminal history of their parents to help predict whether they are likely to recidivate if released pending trial.[93] This might seem like a straightforward case of punishing the child for the sins of a parent. But the relevant point of similarity is not only that Alice and Bob are related, but that *other* parents have tended to pass along these qualities to their children as well. In other words, the Observer concludes that Bob is likely to recidivate if Alice has a criminal history because other people with parents like Alice have also gone on to recidivate.

Similar inferences might follow if Alice and Bob share a household. Recent attempts to predict opioid abuse, for example, operate by assigning risk scores to patients, and then making those scores available to insurers and hospitals to be used in prescription decisions. A higher risk score may be assigned to people whose fellow household members have a history of abuse, even if there is no documented misuse.[94] An Observer predicts that

[91] Joe Pinsker, *People Who Use Firefox or Chrome Are Better Employees*, ATLANTIC (Mar. 16, 2015), https://www.theatlantic.com/business/archive/2015/03/people-who-use-firefox-or-chrome-are-better-employees/387781/ [https://perma.cc/R7H5-JLBR].

[92] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/7UUX-GZGX].

[93] Sample COMPAS Risk Assessment, https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html [https://perma.cc/6G3F-BC6S] (specifically questions 33, 34, 37, and 38).

[94] Mohana Ravindranath, *How Your Health Information is Sold and Turned Into 'Risk Scores,'* POLITICO (Feb. 3, 2019), https://www.politico.com/story/2019/02/03/health-risk-scores-opioid-

Bob is more likely to abuse opioids because it knows that Alice has abused opioids, that Alice lives in the same place as Bob, and that people tend to abuse opioids at a higher rate if they live with other abusers. At first blush, this approach seems to have more in common with the tie-based dependencies that allow an Observer to discover certain things about Bob because it is keeping track of Alice: knowing that Alice has an opioid prescription and that Alice lives with Bob means that an Observer also knows that Bob has easy access to Alice's medicine. Yet this observation alone might not help predict Bob's risk of abuse, unless the Observer has found that other people in similar circumstances as Bob—for instance, those with easy access to others' drugs—have been more likely to abuse opioids.

Information about people's broader social networks can also serve as the basis for inferences about undisclosed or unobserved characteristics.[95] The ability to infer things about people based on their connections to others in a social network first drew significant attention among privacy scholars when researchers demonstrated that sexual orientation could be predicted from one's *friends'* disclosures of sexual orientation.[96] In other words, even if Bob withholds information about his own sexual orientation, his explicit connection to people who *have* disclosed may improve an Observer's ability to accurately infer that of Bob.[97] In addition to sexual orientation, previous research has shown that information from Bob's friends can accurately predict characteristics like religion, location, and who Bob's other friends are—even in the absence of Bob providing such information himself.[98] In fact, even information about Bob's *friends*

abuse-1139978 [https://perma.cc/E63S-HPUR] (noting that risk scores for opioid addiction and overdose take into account "information about a patient's friends, family, and roommates"); *see also* Mathijs de Vaan et al., *Diffusion of Opioids Within the Family Household* (2018) (working paper) (on file with authors).

[95] *See* MacCarthy, *supra* note 20, at 452–53. *See generally* Alan Mislove et al., *You Are Who You Know: Inferring User Profiles in Online Social Networks*, *in* PROCEEDINGS OF THE THIRD ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING 251 (2010).

[96] Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, 14 FIRST MONDAY (2009), https://firstmonday.org/ojs/index.php/fm/article/view/2611/2302 [https://perma.cc/68NA-SE8J].

[97] In fact, researchers have found that a good deal of private information (including sexual orientation) can be inferred about *nonusers* from data given over by users. *See* David Garcia, *Leaking Privacy and Shadow Profiles in Online Social Networks*, 3 SCI. ADVANCES (Aug. 4, 2017), http://advances.sciencemag.org/content/3/8/e1701172.full [https://perma.cc/GHK6-MYMW].

[98] David Garcia, *Privacy Beyond the Individual*, 3 NATURE HUM. BEHAV. 112, 112 (2019). Moreover, even in cases in which Bob does provide some information himself, the predictive power of his friends' data can be even greater than his own. *See id.* ("the data produced by our online friends can be a better predictor of our future behavior than *our own* data" (emphasis added)); James P. Bagrow et al., *Information Flow Reveals Prediction Limits in Online Social Activity*, 3 NATURE HUM. BEHAV. 122, 124–25 (2019) (when predicting the words a person was likely to use in a future social media post, the person's friends' posts were more informative than the person's own past posts).

of friends (that is, people two hops away from Bob in the network) can be the basis for strong prediction of Bob's own attributes.[99]

Recent attempts to leverage social network data in credit scoring rely on inferences like these.[100] These inferences operate in two rather different ways. First, the Observer might believe that social networks exhibit homophily ("love of sameness") as to a particular quality—creditworthiness—and that people in a network are thus likely to be similar to one another in that respect, allowing the Observer to impute what it knows about Bob's associates to Bob himself.[101] Or, the Observer might have learned that *other* people with a similar group of friends as Bob tend to default on their loans, even if Bob's friends do not. In this case, the inference depends on Bob's similarity to *other people with such friends*, not on the similarity between Bob and Bob's friends. Note that both cases involve drawing inferences from past observations. In the first case, the concept of homophily that allows the Observer to judge Bob on the basis of his friends grows out of *prior* observations of social networks exhibiting homophily. In the second case, the Observer may penalize Bob because other people with a similar set of friends as Bob have tended to default in the past.

We may be especially concerned about drawing negative inferences about Bob based on his social ties when those ties are non-volitional: for example, Bob did not choose his parents. It may strike us as deeply unfair to punish Bob for relationships completely outside his control, rather than for his own conduct.[102] But we may still object, in some cases, when we make inferences about Bob based on associates over which he has more control—his friends, for example—based on a concern about chilling his association with potentially "hazardous" Alices. In aggregate, the effect of many Bobs defensively curating their social networks to avoid negative inferences might deepen social stratification, impede socioeconomic mobility, and contribute to polarization, both on- and offline.[103]

[99] Some social networks exhibit monophily (or "love of one")—that is, people may have extreme preferences for others with a particular trait that is not necessarily their own (for example, a woman with friends who are mostly men). Kristen M. Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity Among Friends-of-Friends*, 2 NATURE HUM. BEHAV. 284, 284 (2018). Recent research demonstrates that in networks with this property, Bob's attributes can be strongly predicted by comparing him to *friends of friends*. *Id.* at 284. Altenburger and Ugander's study demonstrates that "friends-of-friends ('the company you're kept in') can disclose private attribute information that is otherwise undisclosed by friends ('the company you keep')." *Id.* at 284.

[100] *See generally* Yanhao Wei et al., *Credit Scoring with Social Network Data*, 35 MARKETING SCI. 234 (2015).

[101] EASLEY & KLEINBERG, *supra* note 29, at 77–79.

[102] *See, e.g.*, *Deuteronomy* 24:16 (New International Version) ("Parents are not to be put to death for their children, nor children put to death for their parents; each will die for their own sin.").

[103] Nizan Geslevich Packin & Yafit Lev-Aretz, *On Social Credit and the Right to Be Unnetworked*, 2 COLUM. BUS. L. REV. 339, 392–99 (2016) ("[R]ational users aware of potential

2. *Socially Salient Group*

As we have argued elsewhere, similar inferences can be drawn about people even without knowledge of their explicit ties to others.[104] Characteristics like gender, race, and age, for example, function as the basis for a whole range of routine inferences, both trivial and profound, even when the people who share these characteristics do not share any direct social connection. Thus, Alice and Bob may be strangers to one another, but their similar age might lead an Observer to believe that Bob shares many of Alice's other observed traits.

Such inferences can be so mundane as to seem unworthy of discussion, but consider their practical import: businesses have long relied on broad demographic categories to differentiate among customers, trusting that such categories reliably predict consumer preferences and behavior, given past observations of people with these characteristics.[105] Knowing only someone's gender, race, or age allows marketers to bring to bear an enormous amount of prior observations about the tendencies of people in these demographic groups. Coveted markets like "men aged 18–24" are a cultural cliché, but the cliché reveals the extent to which we are *always* understood in terms of those like us.[106]

Characteristics like these frequently figure into our everyday inferences because they are highly visible and allow us to quickly bring to bear all that we associate with people with these characteristics. This also gives these characteristics a unique social salience—a deeply felt sense that these are among the most important qualities that define us and that allow us to identify with others like us.

When and why we are likely to view characteristics as *socially salient* characteristics matters for the perceived legitimacy of the resulting inferences. A person might feel social affinity with others on the basis of a wide range of characteristics, ranging from gender to neighborhood or

financial harm from certain online interactions may seek to remove hazardous links while strengthening beneficial social ties. They may sanitize their list of friends by unfriending those who went bankrupt, lost their jobs, live in a poor neighborhood, or are otherwise perceived as financially risky, and by permitting their social network friends to include only those with good careers and financial standing.").

[104] *See, e.g.*, Barocas & Nissenbaum, *supra* note 86, at 62 (explaining that inferences can be drawn about people based on their resemblance to others "with whom they have no meaningful or recognized relations.").

[105] *See generally* JOSEPH TUROW, BREAKING UP AMERICA: ADVERTISERS AND THE NEW MEDIA WORLD (1997).

[106] Of course, decision-makers could also rely on a *lack* of common characteristics, where your dissimilarity from a group about which much is known means that you're assumed not to possess one of the qualities that defines the group. For example, employers might think that certain job applicants could not be qualified because they do not fit the stereotype for the particular role.

occupation, but also seemingly more trivial things like one's food preferences.[107] All of these might be socially salient in the sense that people understand their connections to others in these terms and recognize that these shared traits might imply a broader set of common beliefs and experiences. And yet there are morally relevant differences in how certain characteristics come to be perceived as socially salient. In some cases, certain characteristics might emerge organically as those that feel like an especially relevant point of both social differentiation ("we are a *unique* group!") and social solidarity ("we share so much in *common*!"). In other cases, certain characteristics might have deep social salience because they have served as the basis for imposing the social differentiation necessary for establishing and maintaining social hierarchies. The social significance of a characteristic like race flows in part from the fact that it has served—and continues to serve—as an explicit basis for subjugation.[108]

In recognition of this history, the law forbids inferences on the basis of certain socially salient characteristics in the kinds of high-stakes decisions that shape people's life chances and life course. In particular, discrimination law enumerates a set of characteristics that must not figure into decisions in such areas as employment,[109] credit,[110] and housing.[111] While philosophers and legal scholars have advanced competing normative theories to account for the wrongfulness of discrimination, the law seems to have singled out these characteristics for special treatment because these characteristics have served as the basis for unjust deprivations in the past.[112] Under certain theories, discrimination leads to an unjust deprivation when a decision is driven by a decision-maker's animus or prejudice.[113] Other theories view discrimination's wrongfulness in terms of its coarse groupings and crass stereotypes—inferences that lack sufficient accuracy or precision, thus depriving the deserving of

[107] *See* Kim Severson, *What's for Dinner? The Pollster Wants to Know*, N.Y. TIMES (Apr. 16, 2008), https://www.nytimes.com/2008/04/16/dining/16voters.html [https://perma.cc/636A-VMYU].

[108] *See* Kasper Lippert-Rasmussen, *The Badness of Discrimination*, 9 ETHICAL THEORY & MORAL PRAC. 167, 169 (2006). In the United States, for much of the twentieth century, overt discrimination on the basis of race was commonplace, depriving people of basic rights, and limiting their ability to seek employment, housing, or credit. Public expressions of racial animus were routine; demeaning race-based stereotypes were pervasive. Everyday life was shot through with racism. *See generally* DOUGLAS MASSEY & NANCY DENTON, AMERICAN APARTHEID: SEGREGATION AND THE MAKING OF THE UNDERCLASS (1993).

[109] Civil Rights Act of 1964, Title VII, 42 U.S.C. § 2000e (2018).

[110] Equal Credit Opportunity Act, 15 U.S.C. §§ 1691–1691f (2018).

[111] Fair Housing Act, 42 U.S.C. §§ 3601–3619.

[112] Andrew Altman, *Discrimination*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., 2015), https://plato.stanford.edu/entries/discrimination/ [https://perma.cc/AB8B-LJ98].

[113] *See generally* RONALD DWORKIN, A MATTER OF PRINCIPLE (1985).

important opportunities[114] and needlessly condemning entire social groups to categorical judgment.[115] Still others trace the injustice of discrimination back to the lesser moral status that it seems to accord specific social groups, as if these socially salient characteristics justify showing certain people less respect.[116] The law formalizes these various intuitions by prohibiting certain decision-makers from viewing people as relevantly similar on the basis of these protected characteristics.

Notably, discrimination law prohibits decision-makers from considering these characteristics even when they demonstrate predictive value.[117] Drawing accurate inferences on the basis of membership in a protected class does not make such inferences lawful. Rather, discrimination law forbids decision-makers from falling back on, for example, race- or gender-based heuristics, even when doing so may serve a seemingly rational goal.[118] Instead, the law forces decision-makers to identify other points of similarity among those who hold the sought-after quality.[119] This reflects a belief that people should never be judged merely or even partially on their membership in a protected class because these categories lack *moral* relevance when assessing someone for a job, a loan, or an apartment. In other words, discrimination law views the potential statistical relevance of protected characteristics as itself morally suspect—as a statistical artifact of some past injustice that cannot justify subjecting people to further disadvantage.

While we might resist all sorts of generalizations, profiles, or stereotypes, the law only forbids those that involve socially salient characteristics that are essential to people's self-definition and have served as the basis for systematic oppression in the past.

3. Non-Socially-Salient Group

Finally, statistical analysis may also reveal that certain qualities tend to correlate with characteristics that we might not think of as socially meaningful. Rather than using recognizable demographic categories (like gender, race, or age) to target marketing to a group, advertisers might use behavioral data, like what websites users tend to visit, as the basis for targeting. As Brian d'Alessandro explains, "[t]raditionally, demographic

[114] *See generally* ALAN GOLDMAN, JUSTICE AND REVERSE DISCRIMINATION (2015).

[115] Schauer, *supra* note 82, at 22–24.

[116] *See generally* DEBORAH HELLMAN, WHEN IS DISCRIMINATION WRONG? 35–37 (2008).

[117] *See* Michelle R. Gomez, *The Next Generation of Disparate Treatment: A Merger of Law and Social Science*, 32 REV. LITIG. 553, 562 (2013).

[118] *Id.*

[119] *See id.* at 562–63 n.32.

and lifestyle data has served as a proxy for a good audience. With modern server logs holding behavioral data that tracks every last click, marketing firms can do away with the proxies and build audience segments with a high likelihood to take some sort of specific action."[120] Such innovations mean that we cannot take for granted that the features that define a group will correspond with the existing broad categories of social identity, including the categories protected by discrimination law. These inferences do not rely on an articulable social identity to get at some quality that is difficult to observe directly; rather, they identify entirely new ways to recognize relevant points of similarity. But these are not social groups in any meaningful sense.[121] They do not share any basis for social kinship because the group definition is entirely decoupled from those personal attributes and activities through which we experience identity.

Facebook offers advertisers a targeting mechanism that makes this shift clear. Traditionally, advertisers might have come to Facebook with a pre-established market in mind (e.g., men aged eighteen to twenty-four), which they would target using Facebook's demographic-based tools. In contrast, advertisers can now also ask the social network to find so-called "lookalike audiences"[122]—other users on the social network who resemble those who have previously exhibited an interest in the advertised product or service. Facebook does this by first finding points of similarity among those who have interacted with the advertiser in the past and then looks for *other* users who share these same points of similarity.[123] Notably, Facebook is able to consider an enormous range of possible points of similarity because the social network has such detailed and wide-ranging information about people, their interests, their associations, and their behaviors. As a result, the relevant points of similarity identified by Facebook might involve a mix of characteristics that, if revealed, would not be terribly meaningful to humans.

Even when such groupings demonstrate statistical validity—that is, even when the computationally generated "lookalike audiences" end up exhibiting significantly more interest in the advertised products or services than other users—the characteristics that underlie these groups may feel rather arbitrary. As a consequence, the decision to treat people that belong to this group differently than others can feel equally arbitrary, no matter how accurately group membership predicts a specific outcome of interest. If the characteristics that decision-makers view as relevant do

[120] Brian d'Alessandro, *Actions Predict Louder Than Words*, O'REILLY RADAR (Oct. 23, 2013), http://radar.oreilly.com/2013/10/actions-predict-louder-than-words.html [https://perma.cc/QW69-4LG5].

[121] Vedder, *supra* note 17, at 278.

[122] *Ad Targeting*, FACEBOOK BUS., https://www.facebook.com/business/a/lookalike-audiences (last visited Apr. 13, 2020).

[123] *Create a Lookalike Audience*, FACEBOOK BUS., https://www.facebook.com/business/help/465262276878947?id=401668390442328 (last visited May 9, 2010).

not map onto socially salient differences in the world, the decision-making that relies on this differentiation may lack perceived legitimacy.[124] For example, even though our cultural tastes—in literature, music, film, and cuisine, among other things—often serve as the basis for inferences about our intelligence, we are unlikely to recognize our interest in "Curly Fries" as a reasonable basis for concluding that we must have a high IQ, despite empirical research establishing this correlation.[125] Rather than objecting to the treatment of a specific group, we might object to the idea that this is even a meaningful group in the first place.

The novel computational techniques like machine learning that allow us to identify these new points of relevant similarity raise one of the most basic questions about discrimination law and other frameworks for assessing the fairness of decision-making: what groups are sufficiently vulnerable that they warrant special protection? Even though "[t]he concept of discrimination itself places no substantive restrictions on which salient social groups could, in principle, count for purposes of determining whether an act is an act of discrimination,"[126] new computational methods are deeply troubling because they confound attempts to even arrive at definitions of social groups that hold any resonance.[127] The set of characteristics that discrimination law recognizes as an illegitimate basis for decision-making are those that have served as the basis for subjugation and unjust treatment in the past. The salience of characteristics like gender, race, age, religion, marital status, sexual orientation, and national origin, among others recognized in law, stems from the fact that decision-makers and institutions have previously and explicitly relied on these characteristics to justify their adverse actions. In contrast, we are unlikely to view blue shoe-wearing people as a group entitled to special legal protection under discrimination law, even if critical decisions rest on inferences drawn on that basis. The seemingly random characteristics that support today's machine learning-driven

[124] Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1096–98 (2018).

[125] *See* Michal Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802, 5804 (2013) (finding that one of the Facebook Likes predicting high IQ is "Curly Fries").

[126] Altman, *supra* note 112.

[127] Zarsky raised exactly such objections to an early settlement with DoubleClick that required the company to grant data subjects access to the categories into which they had been slotted, arguing that "the clusters formed are not required to conform to the taxonomy used to divide the population in the past, or answer to catchy names such as 'Pools and Patios' or 'Shotguns and Pickups.' The borders of such classes might be elaborate, not easily defined by a simple category name, and ever-changing." Tal Zarsky, *Cookie Viewers and the Undermining of Data-Mining: A Critical Review of the DoubleClick Settlement*, 2002 STAN. TECH. L. REV. 1, 2 (2002).

inferences may lack the necessary social history to elevate them to the status of legally protected characteristics.[128]

Of course, decisions rendered on the basis of characteristics that lack social salience may still result in disparities along socially salient lines. In many cases, newly identified groups might map closely to traditional social groups because many of the relevant points of similarity will be correlated with protected characteristics. Thus, even when the basis for decision-making lacks the necessary social salience and legal status to bring a charge of disparate treatment, we may be able to observe and contest any resulting disparate impact because we can map these inequalities in outcome back to socially salient and legally protected characteristics.[129]

But in other cases, there will be no apparent mapping between a set of non-socially-salient factors and traditional social groups, taking us out of the familiar realm of discrimination law. In those cases, people subject to adverse decisions have no socially salient criteria from which to make sense of and contest their treatment.[130] As Jonathan Simon concludes, "the effect of actuarial practices is precisely to make it more difficult for groups to intensify their solidarity or to exercise political choice."[131] When inferences are drawn on the basis of broad demographic categories, groups defined by these categories might be able to mobilize on that basis to resist adverse treatment. Responding to these new groupings, in contrast, requires solidarity in the absence of meaningful social ties.[132]

[128] Zarsky suggests a novel approach to this issue: to the extent that predictive analytics premise different treatment on attributes other than those that characterize existing salient social groups—say, shopping habits or dietary preferences—and which people consider central to their self-definition, we might explicitly add such elements as protected categories in antidiscrimination law. Tal Z. Zarsky, *An Analytic Challenge: Discrimination Theory in the Age of Predictive Analytics*, 14 I/S: J.L. & POL'Y INFO. SOC'Y 11, 32–34 (2017).

[129] *See, e.g.*, Muhammad Ali et al., *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, ARXIV (Apr. 19, 2019), https://arxiv.org/abs/1904.02095 [https://perma.cc/SAB5-7UHG].

[130] As de Vries points out, the demise of identity politics has its costs: "awareness of gender biases in language . . . allows me to challenge this structure, for example, by structurally using 'she' and not 'he' as a generic pronoun, or inventing new combinations: s/he. In contrast, in our present post-computational turn era it could easily happen that I'm subjected to profiles, categories and semiotic structures ('suspect type', 'profitable customer,' etc.) of which I am not aware. This lack of knowledge and transparency . . . makes it very difficult to challenge or critique those structures with the tools of the linguistic turn of the 1970s." Katja de Vries, *Privacy, Due Process and the Computational Turn: A Parable and a First Analysis*, *in* PRIVACY, DUE PROCESS AND THE COMPUTATIONAL TURN: THE PHILOSOPHY OF LAW MEETS THE PHILOSOPHY OF TECHNOLOGY 9, 24 (2013).

[131] Jonathan Simon, *The Ideological Effects of Actuarial Practices*, 22 L. & SOC'Y REV. 771, 787 (1988).

[132] Virginia Eubanks notes that digital social sorting prevents the creation of solidarity across race, gender, and class lines, in contrast to physical segregation and containment of people with particular characteristics, which had "the unintentional result" of fomenting such solidarity. VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 184 (2018).

* * *

While the similarity-based dependencies that we have described above all involve the same practical mechanism—recognizing when a person shares the same observable characteristics as those with a sought-after trait and inferring that the person is likely to possess the trait as well—they differ in what they identify as the relevant points of similarity. Being judged on the basis of the company you keep, your membership in a legally protected class, or your seemingly arbitrary grouping with others each raises different concerns. We recoil at the idea of judging a child on the basis of their parents' misdeeds; we reject the notion that employers, lenders, and landlords should be free to discriminate against a person according to their gender; and we fail to recognize our solidarity with others subject to the same treatment.

As a general matter, though, we might argue that people should be judged for what they do, not for the behaviors of those with whom they share certain characteristics. Similarity-based dependencies violate the moral intuition that people deserve to be treated as individuals and subject to individualized judgment. They also deny people the opportunity to abstain from disclosing details that seemingly similar people have willingly divulged. And yet there is no way to avoid using generalizations or avoid being subject to them.[133] As we've pointed out, "[i]nsurance offers the most obvious example of this: the rate that a person pays for car insurance, for instance, is determined by the way other people with similar characteristics happen to drive, even if the person is a better driver than those who resemble him on the statistically pertinent dimensions."[134] One might retort that the insurer could do more to distinguish the person from these drivers, finding additional characteristics that demonstrate that he is a safer driver than most in the group. Yet even this maneuver rests on identifying the ways in which the person is similar to other—safer—drivers; it just does so on a larger number of dimensions. Once the insurer has learned so much about the person that he ceases to resemble anyone that the insurer has seen before, the insurer will not be able to improve the precision of its predictions any further. In this sense, the insurer is simply unable to judge the person as an individual. Any inference about his likely driving must rest on comparisons to others.

Similarity-based dependencies are worrisome when they subject people to coarse generalizations, but they can be equally worrisome when they allow for overly granular distinctions. Once again, insurance nicely illustrates this point. For some, the prospect of being lumped in with

[133] Schauer, *supra* note 82, at 67.

[134] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 688–89 n.67 (2016).

others who are far less healthy may provoke outrage: insurance premiums should reflect differences in people's underlying health risks. For others, the very purpose of insurance is to spread the costs of these risks across policyholders: those in good health should cross-subsidize those in poor health. Some healthcare laws forbid insurers from pursuing too much granularity in pricing premiums.[135] Community rating pricing schemes, for example, require that all people in a given area receive the same price—regardless of health status or other risk factors.[136] The goal of this restriction is to ensure that communities effectively socialize within-community differences in healthcare costs by offering everyone similar premiums, even if insurers could figure out which individuals within the community will prove more or less costly.[137] Doing so prevents insurers from charging high premiums to unhealthy people or denying them coverage altogether.[138]

C. *Difference*

A third sort of dependency involves *difference* between Alice and Bob. In these cases, by revealing some information about *herself* to the Observer, Alice allows the Observer to learn something about *Bob* by making herself distinguishable from Bob. Importantly, Alice need have no connection to, knowledge of, nor similarity to Bob for this to occur—yet Bob's privacy still depends upon Alice's behavior, because it allows the Observer to deduce something about Bob.

Here, we describe three subtypes of this dependency. In *process of elimination*, Alice and Bob are members of a set of people suspected of some proscribed behavior; when Alice tenders information about herself, the Observer can winnow down the set to identify Bob. In *anomaly detection*, the Observer already has information from all parties—but only by comparing Bob's data to that of more "normal" Alices is the Observer able to make meaning from its atypicality. Finally, *adverse inference* occurs when Bob chooses to withhold some bit of information, but most Alices disclose—turning Bob's nondisclosure into a signal that effectively communicates the underlying information to the Observer.

[135] Anthony T. Lo Sasso & Ithai Z. Lurie, *Community Rating and the Market for Private Non-Group Health Insurance*, 93 J. PUB. ECON. 264, 266 (2009).

[136] *Id.*

[137] The Affordable Care Act has an *adjusted* community rating requirement, meaning that insurers are actually allowed to vary individual premiums based on certain factors like age and tobacco use, but not others like previous medical claims. 42 U.S.C. § 300gg(a)(1) (2018).

[138] Santosh Rao, *Q&A: Community Rating & Adjusted Community Rating Under the ACA*, AM. HEALTH LINE, https://www.americanhealthline.com/analysis-and-insight/question-and-answer/q-and-a-community-rating [https://perma.cc/KB64-S8WV].

1. Process of Elimination

In some cases, Alice may be motivated to provide information about herself for exculpatory purposes, which by implication can teach the Observer something about Bob. Assume, for instance, that law enforcement knows that someone using a network in a university dorm made a bomb threat to disrupt a final exam. Assume further that only two people were using that dorm's network at the time of the threat: Alice and Bob. Having narrowed the set of likely suspects to these users, the police approach Alice and interrogate her about her activities at the time the threats were transmitted; Alice, being genuinely innocent and motivated to establish as much to the police, shares with the police a convincing alibi.[139]

Alice's rendering of this information—purely about herself—gives the police information about Bob by implication: specifically, that Bob, as the only other member of the suspect set, is very likely the perpetrator of the bomb threat.[140] In the real world, law enforcement might have access to such a suspect set in the context of "crime-out" investigations, in which police obtain data (like IP addresses, cell phone records, or surveillance camera footage) about all individuals in the vicinity of a crime—often through "reverse search warrants" served on telecoms and other companies, which may provide data on hundreds of devices.[141] They then systematically rule out suspects from this set to identify the perpetrator. As Jane Bambauer notes, "This sort of [third-party] information could give the police an initial suspect pool that could then be winnowed further with the usual detective work."[142]

In this context, Bob's privacy depends on uncertainty, and it is only possible to maintain the uncertainty upon which Bob's privacy depends if

[139] This hypothetical scenario draws from a real example. In 2013, a Harvard student made such a threat. FBI agents were able to identify the student because, though the student had used Tor and an anonymous email program to send the threat, he did so using Harvard's wireless network. Though agents could not tell from network logs precisely what the student was doing using Tor, they *could* tell that he was one of a small set of students using Tor at the time of the threat, which led to his detection. PJ Vogt, *That Bomb-Hoaxing Harvard Student Was Using Tor, But They Caught Him Anyway*, WNYC (Dec. 18, 2013), https://www.wnyc.org/story/harvard-bomb-threat/ [https://perma.cc/9YE4-E4JU].

[140] MacCarthy describes another instance of this dependency in the hypothetical: "Suppose from public data records, it can be inferred that one of two people was involved in a particular transaction and that the person who engaged in the particular transaction was right-handed. If one of the two people discloses that he is left-handed, he or she thereby discloses that the other person engaged in the transaction." MacCarthy, *supra* note 20, at 446.

[141] Jennifer Valentino-DeVries, *Tracking Phones, Google is a Dragnet for the Police*, N.Y. TIMES (Apr. 13, 2019), https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html (calling these "geofence warrants").

[142] Jane Bambauer, *Other People's Papers*, 94 TEX. L. REV. 205, 234 (2015).

certain details about others remain unknown. It may seem unintuitive to think of Bob's interest here as a privacy interest; after all, criminal investigations often require winnowing a suspect pool in this way. But we can just as easily imagine Bob being identified as a political dissident, or whistleblower, or as a person accessing state-censored content, through process of elimination. The point is that the mechanism of learning a secret about Bob relies on the disclosure of Alice and the subsequent deductive reasoning that her disclosure allows.

This situation has three necessary prerequisites.[143] First, Alice and Bob must belong to a closed set of potential targets.[144] Second, the Observer must be able to uniquely distinguish all members of the set, including Bob, relative to one another.[145] And finally, the Observer must be able to learn exculpatory information about all Alices (that is, all members of the set *except* Bob) in order to deduce that Bob is the culprit[146] Pragmatically, this implies that the set of potential targets is of relatively small size to make such evidence-gathering manageable.

Efforts to protect privacy through anonymity generally take one of three approaches to attenuate the conditions required for the Observer to identify Bob—each oriented toward falsifying one of these prerequisites. The first prerequisite—that the suspect set is closed—may be falsified by creating uncertainty about the size of the suspect set. Doing so disrupts the ability of the Observer to bound the number of possible suspects, such that even exculpating everyone except Bob does not conclusively implicate him. In the board game *Clue*, for example, players compete by deducing, via process of elimination, which of six dinner guests at a mysterious estate committed a murder; in every instance of gameplay, the murder was committed by one of these six. But a key twist in the 1985 film adaptation of the board game is that one murder was committed by a character *outside* the suspect set—Mr. Boddy, the owner of the estate, who has been masquerading as the house's butler.[147] The source of surprise for the film's audience is the falsification of the first prerequisite: that only members of the set might be the person of interest. Even reverse search warrants premised on cell phone data are imperfect in this regard: they may not capture data about *all* phones in an area, not to mention potential suspects without cell phones.[148]

[143] *Id.* at 207–8.

[144] *Id.*

[145] *Id.*

[146] *Id.*

[147] CLUE (Paramount Pictures 1985).

[148] Valentino-DeVries, *supra* note 141, (noting that even Google's vast Sensorvault, used to provide geofenced cell phone data to the police, doesn't capture every phone); *see also* Jennifer

These strategies respond to the fact that even when, for example, activity on a network is anonymized, the *use* of that network can often be more readily identified (as in the university bomb threat), giving the Observer a finite number of possible suspects to consider.[149] Certain anonymity-preserving tools therefore aim to facilitate fully unobservable communication such that, to the Observer, the user appears to be making "regular [non-proscribed] network connections, while the user is actually getting connected to destinations that are forbidden by that monitoring entity."[150] In other words, the monitor cannot detect who is using an anonymized network (and is therefore in the suspect set) in the first place, and who is not.

Another approach is to falsify the second prerequisite by making set members incapable of being uniquely identified relative to one another. This is accomplished by creating *plausible deniability* for set members: that is, making set members sufficiently indistinguishable from one another such that no Observer could conclusively determine which one of them committed an observed act, transmitted particular information, or holds a particular identity. One method of doing so involves making everyone in the suspect set look identical. For example, in the "I am Spartacus" technique discussed by Brunton and Nissenbaum,[151] multiple members of the set claim the *same* identity to thwart distinguishability. Alternatively, set members are not made to look identical, but their data are "mixed up" such that it is hard to associate any behavior with a unique

Lynch, *Google's Sensorvault Can Tell Police Where You've Been*, ELECTRONIC FRONTIER FOUND. (Apr. 18, 2019), https://www.eff.org/deeplinks/2019/04/googles-sensorvault-can-tell-police-where-youve-been [https://perma.cc/Q5YP-S678] ("[T]here's a high probability the true perpetrator isn't even included in the data disclosed by Google. For these kinds of warrants, officers are just operating off a hunch that the unknown suspect had a cellphone that generated location data collected by Google. This shouldn't be enough to support probable cause, because it's just as likely that the suspect wasn't carrying an Android phone or using Google apps at the time.").

[149] *See* Vogt, *supra* note 139.

[150] Amir Houmansadr et al., *Cirripede: Circumvention Infrastructure Using Router Redirection with Plausible Deniability*, PROC. 18TH ACM CONF. ON COMPUTER & COMM. SECURITY 187, 188 (2011). A recent Google policy change makes this more difficult by blocking "domain-fronting," a practice in which developers used Google as a proxy to avoid state censorship: "[a]s long as the service was using domain-fronting, all the in-country data requests would appear as if they were headed for Google.com, with encryption preventing censors from digging any deeper." Russell Brandom, *A Google Update Just Created a Big Problem for Anti-Censorship Tools*, VERGE (Apr. 18, 2018), https://www.theverge.com/2018/4/18/17253784/google-domain-fronting-discontinued-signal-tor-vpn [https://perma.cc/SLV4-PW85].

[151] The example derives from the scene in the film *Spartacus* in which "the rebel slaves are asked by Roman soldiers to identify their leader, whom the soldiers intend to crucify. As Spartacus . . . is about to speak, one by one the others around him say 'I am Spartacus!' until the entire crowd is claiming that identity." FINN BRUNTON & HELEN NISSENBAUM, OBFUSCATION: A USER'S GUIDE FOR PRIVACY AND PROTEST 15 (2015); *see also* Zbigniew Kwecka et al., *"I am Spartacus": Privacy Enhancing Technologies, Collaborative Obfuscation and Privacy as a Public Good*, 22 ARTIFICIAL INTELLIGENCE & L. 113, 115–16 (2014).

identity. Obfuscation techniques like swapping loyalty cards to create noise in data about shopping patterns decrease the degree to which observations tied to a particular identifier can be tied to a unique individual.[152] Similarly, someone setting up a Tor exit node maintains plausible deniability as to the transmission of illegal content through the network, as Tor routes many individuals' traffic through multiple servers, making the transmission of any particular piece of information unattributable to a specific person.[153]

A final strategy is to falsify the third prerequisite pragmatically, by expanding the size of the suspect set to create "strength in numbers."[154] Anonymity-preserving networks are more effective at masking individual behavior when they have many users, making identification of a particular user more practically difficult. The effectiveness of this strategy depends in part on the resources that the Observer brings to bear on the task of identifying Bob. As Paul Syverson, one of Tor's creators, noted, having only "a few hundred concurrent users" on an anonymity network "is fine if the goal is simply plausible deniability"—but inadequate for protection against law enforcement or state intelligence, because "[i]f the adversary has the incentives and resources of a nation-state or of organized crime . . . the small anonymity set means that it is now within the resource constraints of the adversary to closely scrutinize the online and offline behavior of everyone identified as participating."[155] However, the size of the set considered "manageable" to sift through has expanded significantly thanks to automated analytic tools available to law enforcement.[156]

One additional point is worth noting. Even when conclusive identification of Bob is impossible—say, because the Observer cannot obtain exculpatory information about *each* Alice, such that it cannot winnow down the set of suspects to a single member—Bob can nonetheless suffer a privacy harm. Say that the Observer knows initially that one member of a set of ten committed a crime. Knowing nothing

[152] BRUNTON & NISSENBAUM, *supra* note 151, at 28–29.

[153] *See Tor Project: Overview*, TOR, https://2019.www.torproject.org/about/overview.html.en [https://perma.cc/RH28-M8V7].

[154] Cooper Quintin, *Tor is for Everyone: Why You Should Use Tor*, GIZMODO (Jun. 15, 2014), https://gizmodo.com/tor-is-for-everyone-why-you-should-use-tor-1591191905 [https://perma.cc/UJL3-M68X].

[155] Paul Syverson, *Practical Vulnerabilities of the Tor Anonymity Network*, *in* ADVANCES IN CYBER SECURITY: TECHNOLOGY, OPERATION, AND EXPERIENCES 60, 65 (D. Frank Hsu & Dorothy Marinucci eds., 2013).

[156] Bambauer, *supra* note 142, at 210 ("Without computers, even the most legitimate searches conducted with a warrant based on probable cause required police to tromp through houses, flip through diaries, and sift through large amounts of personal information unrelated to the investigation. Automated searches, by contrast, can tailor information access so that most irrelevant data is filtered out.").

except the membership of the group, the Observer knows that Bob's likelihood of being the culprit is 10%. But as the Observer rules out Alices from the denominator, Bob's likelihood of being the person within the set who committed the crime increases. Inference need not be definitive to infringe upon privacy.[157] Even if the Observer never identifies Bob, it may deduce that Bob has a high likelihood of having committed the crime, implicating Bob's privacy and potentially leading to him being treated differently.[158]

The use of process-of-elimination techniques can be controversial, especially in law enforcement, because it requires engaging in the sort of broad, unlimited "fishing expedition" that the Fourth Amendment is specifically intended to proscribe. While no court rulings have yet considered the validity of reverse search warrants, privacy advocates and defense attorneys have begun to challenge their constitutionality.[159]

What's more, the existence of process-of-elimination dependencies in investigative contexts reveals a more general and fundamental truth about the relational nature of anonymity. Anonymity is not a characteristic of an individual, nor a result of individual actions that people take to make themselves unidentifiable. Rather, only by preventing an Observer from distinguishing Alice from Bob can Bob maintain some degree of anonymity. What distinguishes us from others is what identifies us. Anonymity, then, depends not only on what is known or unknown about you—but also what is known or unknown about others. Anonymity depends, therefore, on the actions of a collective; it cannot be achieved alone.

The relational nature of anonymity has real consequences for policies intended to protect us from being identified within a group. Laws and policies commonly require that datasets be stripped of personally

[157] *Cf.* MacCarthy, *supra* note 20, at 455 (noting probabilistic nature of information externalities in data mining).

[158] Arvind Narayanan, Joanna Huey, and Ed Felten articulate this principle in the closely related context of dataset re-identification:

> Suppose an analyst can narrow down the possibilities for Alice's record in a de-identified medical database to one of ten records. If all ten records show a diagnosis of liver cancer, the analyst learns that Alice has liver cancer. If nine of the ten show liver cancer, then the analyst can infer that there is a high likelihood of Alice having liver cancer. Either way, Alice's privacy has been impacted, even though no individual database record could be associated with her.

Arvind Narayanan et al., *A Precautionary Approach to Big Data Privacy*, *in* DATA PROTECTION ON THE MOVE 357, 360 (2016); *see also supra* notes 88–89 and accompanying text (discussing inference based on aggregate statistics).

[159] *See, e.g.*, Defendant Okello Chatrie's Motion to Suppress Evidence Obtained from a "Geofence" General Warrant at 1, United States v. Chatrie, No. 3:19-CR-00130-MHL (E.D. Va. Oct. 29, 2019) (moving to suppress evidence of geofence warrant used to obtain cell phone information from all nineteen people in the vicinity of a bank robbery, on grounds of overbreadth and lack of particularity); Lynch, *supra* note 148.

identifiable information (PII). Various laws set out to define the specific pieces of information that count as PII,[160] and to apply special privacy rules when records contain PII.[161] Historically, they have taken the approach of trying to delineate, in advance, the discrete bits of information that would make it possible to associate a record with the person whose information has been captured in the record.[162]

But conceiving of PII as a fixed set of sensitive attributes (for example, name, date of birth, etc.)—as these laws do—makes little sense when any kind of information (for example, movie viewing habits, search queries, etc.) might be uniquely identifying.[163] What policies often overlook is that information is or is not uniquely identifying only when juxtaposed against information about others, regardless of how "sensitive" we might imagine it to be. This insight has turned the notion of PII on its head. What is personally identifiable is not an inherent property of certain pieces of information; it is a function of how effectively any information distinguishes people from one another. Information becomes personally identifiable when it makes someone's records different from everybody else's. A Social Security number, for example, is not identifying in its own right; it is identifying only because no two people have the same number, so it does the work of distinguishing people from one another.[164] Should there be a mistake and two people receive the same number, the number would cease to be uniquely identifying. The relevant lesson is that unique identifiers are necessarily relational: a piece of information about a specific person is only unique—or not unique—when compared to what is known about others.

Various strategies have been developed to deal with this problem.[165] Differential privacy techniques respond to this concern by considering identifiability as a quality of a dataset, rather than tying it to specific identifiers.[166]

[160] For example, HIPAA sets forth eighteen categories of identifiers that must be removed for a dataset to be considered "de-identified" under the rule; these include attributes like names, email addresses, account numbers, and fingerprints.

[161] 45 C.F.R. § 164.514(b) (2019).

[162] *Id.*

[163] Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information,"* 53 COMM. ACM 24 (2010).

[164] *But see* SARAH E. IGO, THE KNOWN CITIZEN: A HISTORY OF PRIVACY IN MODERN AMERICA 64–65 (2018) (discussing the early days of Social Security numbers when they were not unique).

[165] AARON ROTH & MICHAEL KEARNS, THE ETHICAL ALGORITHM (2019) (describing strategies including suppressing certain fields, coarsening the values in each field (for example, move from birth day to birth year), injecting noise into the records (swapping, controlled randomization, etc.), or only reporting aggregate statistics).

[166] *See generally* Cynthia Dwork, *A Firm Foundation for Private Data Analysis*, 54 COMM. ACM 86 (2011) (describing the technical foundations of differential privacy).

2. *Anomaly Detection*

In a related case, many individuals' information is viewable to an Observer, who aggregates the data into a central database. Individuals may be required to render this information by law, or pressured to do so by norm; whatever the cause, the result must be that the Observer has access to data from many parties—including both Alice and Bob. In this case, the Observer seeks to learn whose data are anomalous in the set, in order to detect fraud, wrongdoing, or irregularity meriting investigation or sanction.

Assume, here, that the Observer finds that Bob's behavior is anomalous. The Observer can only render this judgment by establishing what *normal* behavior looks like and assessing abnormal behavior with reference to that norm—a norm established by the aggregated data of multiple Alices. The group of conforming Alices, and the provision of their own data, are essential to constructing the comparison group that defines Bob's data as anomalous. This insight—that identifying deviation from the norm depends, fundamentally, on first identifying the norm—has both statistical and sociological roots. Statistical outliers are defined by their distance from most other observations; the sociology of deviance takes as one of its tenets the fact that "social groups create deviance by making the rules [or norms] whose infraction constitutes deviance, and by applying those rules to particular people and labeling them as outsiders."[167]

Crucially, in this case, the Observer already has Bob's data—but cannot make Bob's data meaningful until they are placed in comparison to others' data. The data at issue may be a single outlying data point or an unusual pattern of behavior, assessed by creating a model from Alice's data patterns. For instance, subpoenas are routinely issued for electricity usage reports from utility companies; unusually high usage as compared to usage by neighbors or by similarly-sized properties can provide incriminating evidence of indoor marijuana grow operations.[168] Bernie Madoff's Ponzi scheme was "made," in part, because the returns on his investments were consistently high in all sorts of conditions, in contrast to the volatility that other investors experienced.[169] Credit card fraud

[167] HOWARD S. BECKER, OUTSIDERS: STUDIES IN THE SOCIOLOGY OF DEVIANCE 9 (1963).

[168] *See, e.g.*, United States v. Gifford, 727 F.3d 92, 101 (1st Cir. 2013) (discussing whether comparator houses for determining abnormal utility usage were sufficiently similar in size); United States v. McIntyre, 646 F.3d 1107, 1112–13 (8th Cir. 2011) (finding no privacy expectation in utility bills used for such purposes).

[169] *Con of the Century*, ECONOMIST (Dec. 8, 2008), https://www.economist.com/node/12818310 [https://perma.cc/292G-9ER2]. Other quantitative indicators of hedge fund fraud include extremely low correlation of returns with those of index funds, and serial correlation in returns over time (i.e.,

detection also depends on spotting deviation from normal spending behaviors, and increasingly uses machine learning methods to do so.[170]

There is one special case of anomaly detection worth noting. In most scenarios, Bob's data are anomalous because they are distinguishable from the norm established by multiple Alices. Sometimes, however, Bob's data are anomalous because of their *in*distinguishability. Plagiarism detection is the clearest example: Turnitin's detection software compares submitted assignments against a database of other submissions and published works, and issues "similarity reports" to alert instructors if a submission too closely resembles another in the set.[171] This case might appear to be the converse of the other instances described, as it is Bob's *lack of* uniqueness compared to Alices' data that is the "tell"—but we can understand it in similar terms. Here, the norm to which most Alices accord is uniqueness, and Bob's deviation from this norm leads to his detection.

No single Alice's contribution to the dataset will "make" Bob in these cases; rather, many Alices' data are required to make Bob's deviation from the norm apparent. Initially, then, this may seem to be a fairly weak form of dependency, as the responsibility for outing Bob is diffuse across many Alices. But this diffusion has two important consequences for privacy. First, it justifies dragnet surveillance by the Observer. Since anomalies can only be detected with reference to norms, *all* Alices' data are required (or so the argument goes) to make meaning of Bob's; an enormous amount of data is collected based on the premise that it is necessary for fraud detection. Former NSA head Keith Alexander famously defended the agency's bulk surveillance program on similar grounds, claiming that "you need the haystack to find the needle."[172] Second, social considerations are unlikely to limit Alice's disclosure: Alice's rendering of her own data likely engenders in her no sense of responsibility for how it contributes to fingering Bob. (Who among us feels that we have helped to call attention to a pot grower through our electricity usage?) This limits the degree to which we can reasonably

abnormally "smooth" returns). In both cases, these results stand out because of their contrast with "normal" behavior. Nicolas P. B. Bollen & Veronika K. Pool, *Suspicious Patterns in Hedge Fund Returns and the Risk of Fraud*, 25 REV. FIN. STUD. 2673, 2677–81 (2012).

[170] *See, e.g.,* Siddhartha Bhattacharyya et al., *Data Mining for Credit Card Fraud: A Comparative Study*, 50 DECISION SUPPORT SYSTEMS. 602 (2011). For a thorough overview of algorithmic methods for anomaly detection across domain areas, see Varun Chandola et al., *Anomaly Detection: A Survey*, 41 ACM COMPUTING SURVEYS 1 (Sept. 2009).

[171] *Similarity Reports: Interpreting a Similarity Report*, TURNITIN, http://turnitin.com/self-service/support-wizard.html#inst-similarity-two [https://perma.cc/8QMJ-WTDD].

[172] Barton Gellman & Ashkan Soltani, *NSA Collects Millions of Email Address Books Globally*, WASH. POST (Oct. 14, 2013), https://www.washingtonpost.com/world/national-security/nsa-collects-millions-of-e-mail-address-books-globally/2013/10/14/8e58b5be-34f9-11e3-80c6-7e6dd8d22d8f_story.html [https://perma.cc/4KYG-MHZF]. For discussion of the haystack metaphor more generally and other instances of its use in surveillance contexts, see Sarah Logan, *The Needle and the Damage Done: Of Haystacks and Anxious Panopticons*, 4 BIG DATA & SOC'Y 1 (2017).

expect Alices to mobilize toward protecting Bob's privacy; if anything, rules and norms counsel Alice *to* disclose.

3. Adverse Inference

In anomaly detection, Bob's deviation from Alices' norm concerns some characteristic of the data tendered by each—an outlying value, say, or an atypical pattern—which arouses the Observer's scrutiny of Bob. But Bob may also deviate by failing to disclose information *at all*, in contravention of Alices' norm to disclose. If normal Alices tend to readily disclose some piece of information to the Observer, and Bob takes some privacy-protective measures to avoid doing so, Bob's very act of impeding the Observer's view may provoke suspicion by being definitionally abnormal. Alices' decision to routinely disclose (or not to routinely withhold) data about themselves can result in Bob's nondisclosure leading to an adverse inference against him, thus implicating his privacy.

We often make inferences from atypical absences of information. If most job applicants detail their employment histories on LinkedIn, *not* having a profile is a red flag to potential employers;[173] if your dating profile lacks a photograph when most have one, prospective partners are likely to infer that you must be unattractive.[174] Nondisclosure is understood by an Observer as a negative signal against the backdrop of other people disclosing. Effectively, the source of the inference is a version of the "nothing-to-hide" fallacy detailed by Daniel Solove: that Bob's unusual, privacy-protective concealment of information signals that he must have some incriminating or undesirable attribute he wishes to keep secret.[175] This is akin to what Julian Sanchez has called *the redactor's dilemma*: the idea that taking uncommon pains to keep certain bits of information hidden is itself an ironically communicative act.[176] For

[173] Allison Cheston, *Recruiters Say: Avoid LinkedIn At Your Peril*, FORBES (May 11, 2012), https://www.forbes.com/sites/work-in-progress/2012/05/11/recruiters-say-avoid-linkedin-at-your-peril/#642d363021f6 [https://perma.cc/9V55-XEUZ] (quoting a staffing agency director saying that "[i]f we are staffing for a recruiting or sales/marketing/business development role, then it is a big red flag if a candidate has either no profile or a limited profile with a low number of connections").

[174] *Cf.* Scott R. Peppet, *Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future*, 105 NW. L. REV. 1153, 1192 (2011) (making a similar observation in the context of selling a car on eBay).

[175] DANIEL J. SOLOVE, NOTHING TO HIDE: THE FALSE TRADEOFF BETWEEN PRIVACY AND SECURITY 22 (2011) [hereinafter SOLOVE, NOTHING TO HIDE].

[176] Julian Sanchez, *The Redactor's Dilemma*, JULIAN SANCHEZ BLOG (Dec. 8, 2009), http://www.juliansanchez.com/2009/12/08/the-redactors-dilemma/ [https://perma.cc/6YLC-Z4XM]. Related phenomena include the negative signals communicated by Glomar responses and Fifth Amendment invocations; though neither legally implicates the speaker for wrongdoing, the public

example, satellite images may blur sensitive military compounds to obscure them from view—which, should nothing else be blurred, effectively calls attention to the location of said compounds.[177] In many cases, the mere use of an encryption or anonymization tool can create suspicion of suspect behavior.[178]

Crucially, the fact that Bob's behavior is *unusual* as compared to the rest of the set—the Alices, who do not take similar privacy-protective measures—makes it communicative. Were it to be common for the Alices to protect their own privacy, it would alleviate the negative signal created by Bob's behavior. Whether or not Bob's behavior is construed as "hiding" depends on the "normal" behaviors of others in a given social context.[179] Wearing clothing at work isn't construed as a concealment, but wearing clothing on the beach might be. Closing the door to the bathroom or having curtains on a house[180] are common privacy-protective measures that arouse no adverse inference because they *are* the norm; we interpret these everyday concealments as polite measures that preserve dignity and decorum.[181]

often perceives them as "saying something by not saying something." Thanks to James Grimmelmann for calling this example to our attention.

[177] Matt Korda, *Widespread Blurring of Satellite Images Reveals Secret Facilities*, FED'N AM. SCIENTISTS (Dec. 10, 2018), https://fas.org/blogs/security/2018/12/widespread-blurring-of-satellite-images-reveals-secret-facilities/ [https://perma.cc/HVU5-K2XW]. Thanks to Arvind Narayanan for calling this example to our attention.

[178] For example, former CIA chief Mike Pompeo wrote in a *Wall Street Journal* op-ed in 2016 that "the use of strong encryption in personal communications may itself be a red flag" for terrorism. Mike Pompeo & David B. Rivkin, Jr., *Time for a Rigorous National Debate About Surveillance*, WALL STREET J. (Jan. 3, 2016, 4:21 PM), https://www.wsj.com/articles/time-for-a-rigorous-national-debate-about-surveillance-1451856106 (last visited Apr. 5, 2020).

[179] Conversely, what behaviors we interpret as privacy-*invasive* are contingent on social norms. The sociologist Erving Goffman coined the term "civil inattention" to refer to the collection of polite behaviors—say, refraining from making sustained eye contact—that people tend to extend to one another in public to maintain feelings of privacy in crowded spaces. ERVING GOFFMAN, RELATIONS IN PUBLIC 385 (1972).

[180] SOLOVE, NOTHING TO HIDE, *supra* note 177, at 23.

[181] Daniel Solove, *A Taxonomy of Privacy*, 154 U. PENN. L. REV. 477, 537 (2006) ("Today's norms and practices . . . call for the concealment of many aspects of the body, bodily functions, and strong displays of emotion. We protect against the exposure of these bodily aspects because this protection safeguards human dignity as defined by modern society . . . . The need for privacy, and therefore the prevention of exposure, is created by the fact that we have social relationships and concomitant norms of dignity and decorum.").

But *unusual* privacy-protective measures—like selective encryption of files[182] or avoiding contact with police[183]—may arouse suspicion precisely because they *violate* perceived norms. As Elizabeth Joh has noted, taking steps to evade surveillance (buying a burner phone, using cash, blocking one's face from cameras, and the like) can raise suspicion, even when such actions are motivated by political or personal privacy preferences, rather than wrongdoing. And they do so because of police's assumptions about which behaviors are commonplace and which are not.[184] As a result, innocent people are incentivized to alter their behavior to avoid being seen as suspicious—since such labeling comes with social costs, even without substantiated wrongdoing.[185] This "tax" on innocent

[182] For example, in *United States v. Cotterman*, 709 F.3d 952 (9th Cir. 2013), border agents in Arizona inspected two laptops belonging to Cotterman, a registered sex offender, and found that certain files on the laptops were password-protected; upon further examination, the agents found child pornography, leading to Cotterman's arrest. In determining whether the border agents had reasonable suspicion to justify the search of Cotterman's computers, the Ninth Circuit considered whether the existence of specific password-protected files was a factor supporting the reasonableness of the search. *Id.* at 969. In so doing, the court grappled explicitly with how *normal* such a privacy-protective measure is, noting that "[w]e are reluctant to place much weight on [the password-protection of the files] because it is commonplace for business travelers, casual computer users, students and others to password protect their files. Law enforcement 'cannot rely *solely* on factors that would apply to many law-abiding citizens,' . . . and password-protection is ubiquitous." *Id.* (emphasis added, internal citation omitted). Moreover, password-protection of an entire *device*, rather than files within a device, would not support a search, because device-level password-protection is an extremely common means of ensuring security. *Id.* n.17. The court ultimately held that selective password-protection of files *could* help support a border search when combined with other factors creating reasonable suspicion.

[183] Illinois v. Wardlow, 528 U.S. 119 (2000). In *Wardlow*, police were deemed to have had reasonable suspicion to stop an individual based on his "unprovoked flight upon noticing" them in a high-crime area. *Id.* at 124. In reaching this conclusion, the Court reasoned that "[h]eadlong flight . . . is the consummate act of evasion: It is not necessarily indicative of wrongdoing, but it is certainly suggestive of such." *Id.* In a partial dissent, Justice Stevens called attention to several alternative rationales that would weaken the signal of suspicion conveyed by flight, noting that "[a] pedestrian may break into a run for a variety of reasons—to catch up with a friend a block or two away, to seek shelter from an impending storm, to arrive at a bus stop before the bus leaves, to get home in time for dinner, to resume jogging after a pause for rest, to avoid contact with a bore or a bully, or simply to answer the call of nature—any of which might coincide with the arrival of an officer in the vicinity." *Id.* at 128–29. He added that unprovoked flight from police "is neither 'aberrant' nor 'abnormal'" for members of minority groups for whom contact with police may itself be dangerous. *Id.* at 133.

[184] Elizabeth E. Joh, *Privacy Protests: Surveillance Evasion and Fourth Amendment Suspicion*, 55 ARIZ. L. REV. 997, 1016–17 (2013) ("Because suspicious behavior is often unusual behavior, police judgments about criminally suspicious behavior are necessarily hunches about abnormality, regularity, and conformity.").

[185] *See* L. Rush Atkinson, *The Bilateral Fourth Amendment and the Duties of Law-Abiding Persons*, 99 GEO. L.J. 1517 (2011). Atkinson points out that the Fourth Amendment's protection against only *unreasonable* searches incentivizes even law abiders to act non-suspiciously. *Id*. at 1520. Because people have no legal recourse against a reasonable but erroneous search, they internalize the private costs (hassle, loss of dignity, reputational harm, etc.) of the risks of being searched. *Id.* at 1543. Atkinson argues that this aspect of the Fourth Amendment channels law abiders to act in a way that minimizes erroneous searches by making it easier for police to identify wrongdoers by their suspicious activity. *Id.* at 1520. In other words, it incentivizes people to act "normally."

privacy-protective activity is imposed more heavily on groups that are over-policed or otherwise subject to heightened suspicion.[186] When privacy-protective behavior intersects with constitutional rights—for example, when the behavior consists of refusing a warrantless search of one's person or property—courts have struggled with the degree to which such behavior can be interpreted as a signal of wrongdoing.[187]

The line between a negative signal and a social norm is a malleable one. Should others begin to withhold information in the same manner as Bob, the adverse inference created by Bob's concealment is weakened. This argument counsels in favor of understanding privacy as a collective good that can be collectively protected. We previously discussed why tactics like obfuscation and the use of anonymity networks attenuate the conditions required for identifying Bob through process of elimination;[188] at the same time, when these strategies become more common, they normalize privacy-protective activity and weaken the signal of wrongdoing that may be associated with it. As Bruce Schneier writes:

> Encryption should be enabled for everything by default, not a feature you turn on only if you're doing something you consider worth protecting . . . . If we only use encryption when we're working with important data, then encryption signals that data's importance. If only dissidents use encryption in a country, that country's authorities have an easy way of identifying them. But if everyone uses it all of the time, encryption ceases to be a signal. No one can distinguish simple chatting from deeply private conversation. The government can't tell the dissidents from the rest of the population. Every time you use encryption, you're protecting someone who needs to use it to stay alive.[189]

[186] *Id.* at 1524–25.

[187] Generally, the invocation of Fourth and Fifth Amendment rights cannot be considered as evidence of guilt, as doing so places "an unfair and impermissible burden" on the exercise of such rights, robbing them of their meaning. United States v. Prescott, 581 F.2d 1343, 1351 (9th Cir. 1978). However, some state statutes *do* allow a driver's refusal to submit to blood and breath tests to be considered as evidence of wrongdoing on the basis of implied consent laws that attach to operation of a vehicle in those states. The tension between these evidentiary consequences and Fourth Amendment protection is currently unresolved. *See* Petition for Writ of Certiorari, Bell v. Pennsylvania, __ U.S. __ (No. 19-622), https://www.supremecourt.gov/DocketPDF/19/19-622/122507/20191114190036011_No.__PetitionForAWritOfCertiorari.pdf [https://perma.cc/WTA8-J3K2] (seeking certiorari to determine evidentiary consequences of refusal to submit to warrantless blood test). *See generally* Kylie Fisher, *Save Your Breath: A Constitutional Analysis of the Criminal Penalties for Refusing Breathalyzer Tests in the Wake of* Birchfield v. North Dakota, 94 WASH. L. REV. ONLINE 1 (2019).

[188] *See supra* section II.C.1.

[189] Bruce Schneier, *Why We Encrypt*, SCHNEIER ON SECURITY (June 23, 2015), https://www.schneier.com/blog/archives/2015/06/why_we_encrypt.html [https://perma.cc/VL6Z-C236]; *see also* Joh, *supra* note 169, at 1004 (describing everyday efforts to thwart surveillance as being a form of "privacy protest" with social value); Scott Skinner-Thompson, *Performative Privacy*, 50 U.C. DAVIS L. REV. 1673 (2016) (characterizing same as expressive acts of resistance).

Making Bob's privacy dependent on the collective action of many Alices is a risky proposition, since—as we have described—Alice may be disincentivized from taking privacy-protective steps herself. One of the most powerful structural ways to normalize these steps is to obviate the collective action problem through design defaults.[190] Design defaults effectively scrub privacy-protective behaviors of meaningful signals by making such behaviors both widespread and deniable.[191]

The Observer can make adverse inferences from Bob's nondisclosure when "normal" Alices disclose. In addition, a special case of adverse inference from nondisclosure, *unraveling*, emerges when Alice and Bob are explicitly ranked against one another, competing for some service or position. The logic that underlies unraveling is this:[192] disclosure of some type of information is officially voluntary, but the Observer attaches value to disclosure with some incentive or benefit—say, a discount—if the discloser has some desirable attribute. Assume that Alice initially discloses some attribute that puts her in a good light, and that Bob has a comparatively weaker value for that attribute. Upon Alice's disclosure, Bob can be affected in one of two ways. He might disclose, despite preferring not to, in order to distinguish himself from even worse performers. Or he might *not* disclose and, as a result, be subject to adverse inference based on his nondisclosure. Though Bob technically has the power to choose, his choice is illusory: whether or not he discloses, Alice's provision of information facilitates the Observer learning more about Bob *regardless* of what Bob does, either explicitly, should Bob disclose, or through adverse inference, should he withhold. The mechanism that drives unraveling is the incentivized desire to distinguish oneself from a group—specifically, to designate oneself as above average on some dimension. But as more people disclose, the average of the remaining non-disclosing pool shifts ever downward, until everyone has

[190] Ironically, market dominance and service lock-in have fortunate side effects when their default settings are privacy-protective. Dieter Bohn, *Why I Turned On iMessage*, VERGE (Jun. 13, 2019), https://www.theverge.com/2019/6/13/18677644/imessage-iphone-apple-secure-encrypted-chat-moral-imperative-signal-rcs-hangouts [https://perma.cc/4KKT-PFXR] (noting the benefits of iMessage being encrypted by default rather than trying to convince people to switch to a third-party encrypted chat app).

[191] Karen Levy & Bruce Schneier, *Privacy Threats in Intimate Relationships*, J. CYBERSECURITY (forthcoming 2020) (on file with authors). Since many people never adjust privacy settings, making data private by default increases the number of people using them, and also obviates the need for a person to take affirmative steps to protect privacy, which can function as a "tell."

[192] The most thorough treatment of unraveling in the information privacy law context is Peppet, *supra* note 174; *see also* MacCarthy, *supra* note 20, at 26–27 (summarizing adverse inferences to be drawn from nondisclosure in eligibility decisions).

either disclosed or is subject to inference that their value for the attribute is of the lowest quality.[193]

To illustrate, imagine that a university that used to require standardized test scores for admissions decisions opts to make such disclosure voluntary.[194] Alice, who has a very high score, will surely share this information with the university. Bob, who has a somewhat lower but still respectable score, will be incentivized to disclose as well, because he wants to avoid being lumped in with the masses; he wants the Observer to know that he is superior (along this dimension) to students with even lower scores. If Bob refrains from sharing his score, the university may infer that his nondisclosure masks poor performance: if he did well, why not share? Bob's disclosure has the same effect on someone with a marginally lower score than Bob—and on down the line—such that in the end, nearly everyone has disclosed, and those who haven't are assumed to be the worst of the lot.

The configuration shares some resemblance with identification through process of elimination,[195] in which Alice's disclosure is also self-interested. There, disclosure serves to exculpate Alice from culpability, increasing Bob's probability of being identified. Here, Alice's disclosure favorably differentiates her from the crowd—and in so doing, impugns Bob if he does not also disclose. Both Alice's and Bob's disclosures are purportedly individual and voluntary—but they really aren't, precisely because Alice and Bob are in competition with one another for some sort of favorable treatment. Disclosure is a way to convince the observer that "I am not like these other people; I am better than (at least some of) them," and its effect is to create social cleavage and undermine solidarity.[196]

As Scott Peppet describes, unraveling has limits: it only applies when disclosure is low-cost and credible, and when no countervailing norms militate against it.[197] Legal constraints attempt to prevent unraveling in a variety of ways. They may prohibit (even voluntary) disclosures, but doing so is difficult to justify and often in conflict with First Amendment interests—and self-interested actors can often make end-runs around such rules though signaling in other ways.[198] In the alternative, law might

[193] *See* Peppet, *supra* note 174, at 1181.

[194] *See id.* at 1196. In this hypothetical, the university has historical data about applicant scores, owing to its recent policy change. In the absence of such data, the university would have to either ascribe some central score to everyone in the pool or use readily available data as proxies. *Id.* at 1161.

[195] *See supra* section II.C.1.

[196] *See* Peppet, *supra* note 174, at 1202 ("The ability to disclose—even at the risk of unraveling privacy—brings with it the ability to seek economic advantage. There are distributive stakes here.").

[197] *See id.* at 1190–96.

[198] *See id.* at 1198–99. For a review of other disclosure and nondisclosure rules in law, see Adam M. Samaha & Lior J. Strahilevitz, *Don't Ask, Must Tell—And Other Combinations*, 103 CALIF. L. REV. 919 (2015).

prohibit the use of certain forms of data in making some decision, such that disclosure ceases to accomplish its intended end. We apply such rules in the Fair Credit Reporting Act (which restricts the bases on which creditors can deny credit) and in health care statutes that limit what data insurers can use in setting premiums.[199] But in many contexts, such policies are likely to face opposition from those who would benefit from disclosure, and as such it may be very difficult to mobilize political support for them.

* * *

Each of the difference-based dependencies described in this section implicates concerns about how we stand out from the crowd. There's little Bob can do to protect his privacy in these cases; in some cases, any attempts he might make to do so may, perversely, make him stand out even more. The collectivity is essential to privacy preservation here. But it can be very difficult to muster collective will among many Alices to help Bob achieve privacy, since they are likely either unaware of the effects of their disclosures or acting out of requirement or self-interest.

Because many individuals' data are required to pinpoint a suspect through process of elimination, to identify anomalous data, or to make meaning of abnormal nondisclosure, the Observer's techniques tend toward the mass collection of information via dragnets or strong expectations of widespread disclosure (enforced through rule, norm, or incentive). Given the difficulties of mobilizing collective action, difference-based dependencies are best addressed through restrictions on this mass collection—for example, prohibiting reverse search warrants, barring institutions from using certain types of data in making decisions, or leveraging design to impede the signals sent by proactive privacy protection.

III. CASE STUDY: GENETIC DEPENDENCIES

We turn here to examination of a specific context in which multiple forms of dependency are at work. We do this to demonstrate the value of clarifying precisely what dependencies are at stake in a given situation, and what technical and legal tools are available to protect the interests that each implicates. Of late, and due in large part to the proliferation of genealogical databases and DNA home-testing kits,[200] legal scholars and the public are becoming more attuned to the privacy implications of

[199] *See* Peppet, *supra* note 174, at 1199–200 (discussing these and other examples, and also noting the limitations of such policies).

[200] *See generally* ALONDRA NELSON, THE SOCIAL LIFE OF DNA: RACE, REPARATIONS, AND RECONCILIATION AFTER THE GENOME (2016) (discussing the social ramifications of increased access to genetic testing, particularly with respect to race).

genetic data.[201] Genetic information is widely seen as worthy of special attention and protection. DNA has a set of qualities that give it particular value for a wide range of investigative and predictive uses: it is uniquely identifying; it demonstrates immutable relationships with others;[202] and it predicts propensities for future risks. These qualities also make it readily exploitable for all three forms of dependency we identify: tie-based, similarity-based, and difference-based. The genetic context presents a telling example of how different configurations of privacy dependencies can become entwined in practice, and the implications of those entanglements for privacy regulation.

A. Tie-Based

The 2018 capture of the Golden State Killer[203] aroused public attention about the forensic power of genetic data. A good deal of the investigation of the Golden State Killer relied on tie-based dependencies—genetic connections revealed in genealogy databases. The investigation initially identified the great-great-great-grandparents of Joseph DeAngelo; from there, investigators relied on detective work, constructing family trees with thousands of relatives to develop a suspect set.[204] As discussed in section II.B, other forms of dependency were also exploited in the investigation.[205]

More and more people are subject to identification based on familial ties—without their own direct participation or opportunity to withhold consent—as home genetic testing websites grow and as law enforcement increasingly gains access to them. Researchers estimate that about 60% of European-descended Americans can be genetically linked to at least a third cousin through commercial DNA testing services to which law enforcement has access, based on the fact that those databases currently include samples from about 0.5% of the U.S. adult population.[206] If the

[201] *See* Jake Weidman et al., *On Sharing Intentions, and Personal and Interdependent Privacy Considerations for Genetic Data: A Vignette Study*, 16 IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY & BIOINFORMATICS 1 (2019) (examining what factors affect the likelihood of sharing genetic data with different organizations).

[202] Ram, *supra* note 77, at 877.

[203] *See supra* note 62 and accompanying text.

[204] Justin Jouvenal, *To Find Golden State Killer, Investigators First Found His Great-Great-Great Grandparents*, WASH. POST (Apr. 30, 2018), https://www.washingtonpost.com/local/public-safety/to-find-alleged-golden-state-killer-investigators-first-found-his-great-great-great-grandparents/2018/04/30/3c865fe7-dfcc-4a0e-b6b2-0bec548d501f_story.html (last visited Apr. 17, 2020).

[205] *See infra* section II.B.

[206] Yaniv Erlich et al., *Identity Inference of Genomic Data Using Long-Range Familial Searches*, 362 SCIENCE 690 (2018); Jocelyn Kaiser, *We Will Find You: DNA Search Used to Nab Golden State Killer Can Home in on About 60% of White Americans*, SCI. MAG. BLOG (Oct. 11, 2018, 2:00 PM),

participant figure rises to 2%, expected to occur within two to three years, 90% of the European-descended population of the United States will be identifiable by a third cousin or closer.[207] The disclosure decisions of the few effectively determine the privacy outcomes for the many.

Several signs indicate that the exploitation of tie-based genetic dependencies is poised to become even more widespread. Police have begun using the technique more frequently, and to investigate less serious crimes. In March of 2019, police in Utah cooperated with genealogy database GEDmatch to identify a suspect in the assault of a church organist. Despite GEDmatch's terms of service specifying that it would only cooperate with law enforcement for homicide and sexual assault cases, it broke its own rules by pointing police to the seventeen-year-old perpetrator's great-uncle, whose DNA was in its database.[208]

Further, some genealogy companies have shown a marked change in how they talk about the use of their databases for investigative purposes. FamilyTreeDNA, which has a partnership with the FBI, implores customers in a new ad campaign to help "provide the missing link"[209] through DNA samples that could help solve violent crimes. FamilyTreeDNA's founder said the company had "a moral responsibility" to help solve cold cases and bring families closure.[210] The genomics company Verogen, upon its purchase of GEDMatch, made a similar appeal: "Never before have we as a society had the opportunity to serve as a molecular eyewitness, enabling law enforcement to solve violent crimes efficiently and with certainty."[211]These claims do not, however, specify the mechanism through which customers can (in

https://www.sciencemag.org/news/2018/10/we-will-find-you-dna-search-used-nab-golden-state-killer-can-home-about-60-white [https://perma.cc/ETF4-BVP4].

[207] Heather Murphy, *Most White Americans' DNA Can Be Identified Through Genealogy Databases*, N.Y. TIMES (Oct. 11, 2018), https://www.nytimes.com/2018/10/11/science/science-genetic-genealogy-study.html [https://perma.cc/SDT2-YJW2].

[208] Peter Aldhous, *The Arrest of a Teen on an Assault Charge Has Sparked New Privacy Fears About DNA Sleuthing*, BUZZFEED NEWS (May 14, 2019), https://www.buzzfeednews.com/article/peteraldhous/genetic-genealogy-parabon-gedmatch-assault [https://perma.cc/Q7SQ-R7BU].

[209] Jennings Brown, *Ancestry-Testing Company: It's Our 'Moral Responsibility' to Give the FBI Access to Your DNA*, GIZMODO (Apr. 3, 2019, 11:15 AM), https://gizmodo.com/ancestry-testing-company-it-s-our-moral-responsibilit-1833774781 [https://perma.cc/JN5F-23QL].

[210] *Id.* At the same time, some companies have restricted law enforcement access in response to privacy concerns. Kristen V. Brown, *DNA Site that Helps Cold-Case Sleuths Curbs Access for Cops*, BLOOMBERG (June 10, 2019, 7:00 AM), https://www.bloomberg.com/news/articles/2019-06-10/dna-site-that-helps-cold-case-sleuths-curbs-access-for-police [https://perma.cc/636A-VMYU].

[211] Julian Husbands, *GEDMatch Partners with Genomics Firm*, VEROGEN (Dec. 9, 2019), https://verogen.com/gedmatch-partners-with-genomics-firm/ [https://perma.cc/K525-P4DU].

FamilyTreeDNA's words) "crowd-source crime solving"[212]—by providing incriminating information about their own family members.

B. Similarity-Based

Genetic information is often the basis for similarity-based dependency, most commonly in the context of inferred characteristics based on genetic markers. Even without understanding the biological pathways between genes and propensities for disease, researchers can make inferences about a person's probability of developing various health conditions based on their resemblance to others with similar genetic profiles. Pharmaceutical giant GlaxoSmithKline recently purchased a $300 million stake in home DNA testing company 23andMe that brings with it exclusive rights to use 23andMe's trove of data to develop drug targets.[213] In addition to having customers' genetic information, 23andMe sends regular surveys to its customers to capture phenotypic and behavioral data,[214] and runs a health hub "where customers can share information about how they manage 18 common health conditions"—giving the company and its partners access to self-reported information on condition prevalence and efficacy of various treatments for people with known genetic profiles.[215] Glaxo's hope is to generalize from 23andMe's data by making inferences about the predispositions and treatment responsiveness of potential customers who *don't* have 23andMe profiles, but share genetic patterns with those on the platform.

Notably, disease risk scoring using genetic data is far more accurate for European-descended individuals than for those with African, Latino, or Asian ancestries—a consequence of the Eurocentric composition of individuals who have participated in scientific genome studies. This underrepresentation—which has been called "the major ethical and scientific challenge surrounding clinical translation and, at present, the most critical limitation to genetics in precision medicine"[216]—stands in

[212] Brown, *supra* note 210.

[213] Megan Molteni, *23andMe's Pharma Deals Have Been the Plan All Along*, WIRED (Aug. 3, 2018, 3:28 PM), https://www.wired.com/story/23andme-glaxosmithkline-pharma-deal/ [https://perma.cc/NBD8-6SMP].

[214] *Id.*

[215] Megan Molteni, *23andMe Wants You to Share Even More Health Data*, WIRED (Apr. 20, 2018, 7:00 AM), https://www.wired.com/story/23andme-wants-you-to-share-even-more-health-data/ [https://perma.cc/8XUG-RQP9].

[216] Alicia R. Martin et al., *Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities*, 51 NATURE GENETICS 584, 584 (2019).

stark contrast to the *over*representation of those groups in DNA databases used for criminal investigation.[217]

Insurers may also be interested in making inferences about a person's future health based on genetic profiles.[218] The Genetic Information Nondiscrimination Act (GINA) bans health insurers from using genetic information about an individual or that individual's family members to make coverage determinations or set premiums—but long-term care insurers, disability insurers, and life insurers are not restricted from doing so.[219] The life insurer YouSurance, for example, uses epigenetic data—information about gene expression modified by influences from the environment and behaviors—to set differential rates for its policies.[220] As scientific understanding of genes and their expression continues to progress—and as more people participate in scientific studies and exchange information on their health conditions on 23andMe—the basis for genetic inference and the predictive power of these data will only grow.

Similarity-based dependencies that rely on phenotype are also implicated in some criminal investigations. Forensic DNA phenotyping is used to predict the physical traits of an unknown person who has left a DNA sample at a crime scene, based on probabilistic associations with traits in other people with similar DNA profiles. The technique can be used to predict traits like hair color, skin color, eye color, freckling, height, baldness, and earlobe attachment.[221] In several cold cases, law enforcement agencies have released DNA phenotype composite images in hopes of identifying a suspect.[222] Sometimes, dependencies are utilized in tandem: in the Golden State Killer case, after several men were identified as persons of interest based on familial matching and circumstantial evidence (like residency in California during the time of the murders), genetic genealogists used health risk and eye color analysis websites to determine that people with the genetic profile of the unknown

[217] Erin Murphy & Jun Tong, *The Racial Composition of Forensic DNA Databases*, 108 CALIF. L. REV. (2020) (forthcoming).

[218] *See supra* notes 136–138 and accompanying text.

[219] Michelle Andrews, *Genetic Tests Can Hurt Your Chances of Getting Some Types of Insurance*, NPR (Aug. 7, 2018, 9:00 AM), https://www.npr.org/sections/health-shots/2018/08/07/636026264/genetic-tests-can-hurt-your-chances-of-getting-some-types-of-insurance [https://perma.cc/QFH8-MERL].

[220] YOUSURANCE, https://www.yousurance.com/science/ [https://perma.cc/4F9U-W6GP].

[221] *See* Charles E. MacLean & Adam Lamparello, *Forensic DNA Phenotyping in Criminal Investigations and Criminal Courts: Assessing and Mitigating the Dilemmas Inherent in the Science*, 8 RECENT ADVANCES IN DNA & GENE SEQUENCES 104, 104 (2014).

[222] *See, e.g.,* Sean Alloca, *First DNA-Phenotyped Image of 'Person of Interest' in Double Homicide*, FORENSIC MAG., Jan. 15, 2015 (describing release of phenotype composite image in search for suspect in South Carolina homicide).

sample were likely to have blue eyes and to bald prematurely. Only one of the identified suspects from that set—Joseph DeAngelo—had those characteristics.[223]

C. Difference-Based

Genetic data can also be the basis of the third type of privacy dependency, based on difference between people. As discussed, genetic data is commonly used as a way to identify an unknown person—for instance, by matching a DNA sample from a crime scene with information from genetic profiles of known individuals (e.g., against a database of samples held by law enforcement or a genetic testing company). In other cases, though, investigators might try to collect DNA samples from all members of a circumscribed set of suspects, on the belief that testing an unidentified sample from a crime scene against these newly collected samples will either identify the culprit or rule out the innocent. These cases meet the conditions in which *process of elimination* dependencies can arise:[224] there is an (assumed) finite set of individuals who may have committed the offense, the individuals are genetically distinguishable as to one another, and the suspect set is not so large as to make bulk collection pragmatically impossible. When these conditions are met, genetic data can be used to identify Bob—who may not tender a sample—because innocent Alices submit their own, winnowing down the suspect set. Unlike the tie-based genetic dependencies in familial searches of existing genetic databases, collecting *new* DNA samples creates a difference-based genetic dependency.

Dragnet DNA sweeps involve the mass collection of genetic samples for purposes of identifying a suspect. The technique is often used in cases in which the suspected perpetrator of some wrong is assumed to be one of a manageably sized finite set.[225] The collection may be premised on rule or on norm. In the former case, a search warrant or organizational rule may mandate the mass collection of samples. For example, in 2018, a woman in an Arizona long-term care facility who had been severely incapacitated for a decade gave birth to a baby, leading to a strong presumption that she had been sexually assaulted by a caretaker during that period. Police served a search warrant on the facility seeking DNA

[223] Heather Murphy, *She Helped Crack the Golden State Killer Case. Here's What She's Going to Do Next.*, N.Y. TIMES (Aug. 29, 2018), https://www.nytimes.com/2018/08/29/science/barbara-rae-venter-gsk.html [https://perma.cc/E62D-3MGS].

[224] *See supra* section II.C.1.

[225] David M. Halbfinger, *Police Dragnets For DNA Tests Draw Criticism*, N.Y. TIMES (Jan. 4, 2003), https://www.nytimes.com/2003/01/04/us/police-dragnets-for-dna-tests-draw-criticism.html [https://perma.cc/79YS-D9NS] (describing several DNA dragnets used by law enforcement, and noting that such mass screenings have been more successful "when the police have narrowed their focus to smaller groups").

samples from *all* its male employees[226]—under the assumption that one of them was very likely to have raped the woman—in order to determine whose DNA matched that of the baby.[227]

In other cases, individuals may be coerced into giving samples based on a desire to self-exculpate, a belief that they have no right to refuse, or based on some more generalized sense of civic duty—knowing that by tendering their own DNA, they are not only reducing the size of the candidate pool, but are also helping to create a norm toward disclosure, the violation of which will create *adverse inference* against the non-compliant.[228] In a 2003 case, for example, police asked 800 Louisiana men for DNA to be matched against unidentified murder scene samples; one man (who was not implicated in the killings) was told by police that submitting to a cheek swab "was his choice . . . but if he refused, [the police] would get a court order that would get in the newspapers and then everyone would know he was not cooperating."[229] In a similar Massachusetts sweep in 2005 in which *all* adult males in a town were asked to submit DNA, one man who volunteered his sample said he did so because "[i]f it gives them one less suspect, that's fine by me . . . . I don't have anything to hide."[230] Similar norms were invoked in response to a 2004 Oklahoma dragnet: police announced that failure to cooperate by *voluntarily* submitting a sample "leaves an open end out there for us to look at."[231]

In practice, these investigations often implicate multiple privacy dependencies. An investigation of the 1998 murder of an eleven-year-old boy in the Netherlands, for example, relied on a combination of familial matching and adverse inference by nondisclosure: in seeking to identify DNA found on the victim's clothing, prosecutors sought voluntary collection from over 20,000 Dutchmen based on familial matching, as

[226] *DNA Samples Sought at Facility Where Woman in Vegetative State Gave Birth*, CBS NEWS (Jan. 9, 2019), https://www.cbsnews.com/news/dna-samples-sought-at-facility-where-woman-in-vegetative-state-gave-birth/ [https://perma.cc/Z3U3-NZKU].

[227] This case is unusual in that investigators' search is premised not on a DNA sample from the culprit himself, only the baby with whom he must have a partial genetic match. In this sense, the investigative strategy is also taking advantage of tie-based genetic dependencies.

[228] For discussion of potential Fourth and Fifth Amendment challenges to "voluntary" DNA dragnets, see Lauren Kirchner, *DNA Dragnet: In Some Cities, Police Go From Stop-and-Frisk to Stop-and-Spit*, PROPUBLICA (Sept. 12, 2016), https://www.propublica.org/article/dna-dragnet-in-some-cities-police-go-from-stop-and-frisk-to-stop-and-spit [https://perma.cc/J6XT-LWWG].

[229] *Halbfinger*, *supra* note 225.

[230] Jonathan Finer, *Baffled Police Try DNA Sweep*, WASH. POST (Jan. 12, 2005), http://www.washingtonpost.com/wp-dyn/articles/A2097-2005Jan11.html [https://perma.cc/KVB5-9XND].

[231] Rebecca Leung, *DNA Dragnet*, CBS NEWS (Sept. 10, 2004), https://www.cbsnews.com/news/dna-dragnet/ [https://perma.cc/7FVG-MLPL].

well as mandatory collection from 1,500 men "of special interest" in the case.[232] In the end, a match was located based on a combination of strategies: when one man failed to submit an obligatory sample and could not be located by police, police took DNA samples from his family members, which matched the crime scene DNA.[233] And even when only one dependency is exploited for a particular case, samples may be *retained* for use in other cases in which other dependencies are at play. For instance, after police knocked on the doors of over 500 residents of East New York, asking them to submit "voluntary" cheek swabs to eliminate them from the suspect pool for a nearby murder, the DNA of even those *exculpated* in the current investigation was retained for future investigations, including familial matching.[234]

D. The Law of Genetic Dependencies

Consider, then, the variety of ways privacy dependencies attach to the use of genetic data. All three types of dependency have been harnessed in criminal investigations—and often in combination with one another, exploiting the capabilities of each. As we shall see, all three types are surprisingly implicated by employment nondiscrimination law as well.

In 2015, supervisors in an Atlanta grocery warehouse were frustrated by repeatedly finding piles of human feces on the floor of their facility. The supervisors made a list of potential suspects in an attempt to pinpoint the "devious defecator,"[235] based on which of their employees' work schedules seemed to align with the timing and location of the offenses; they then asked two employees from that list, whom they suspected of having left the piles, to give DNA samples to be matched against the feces.[236] The workers gave the samples, saying they feared for their jobs should they refuse. As it turned out, neither was a match. They subsequently sued their employer under the Genetic Information Nondiscrimination Act (GINA).[237]

[232] Milan Schreuer, *17,500 Dutchmen Gave Their DNA in a Murder Inquiry. After 20 Years, an Arrest.*, N.Y. TIMES (Aug. 27, 2018), https://www.nytimes.com/2018/08/27/world/europe/netherlands-murder-dna.html [https://perma.cc/4MJN-TUXP].

[233] *Id.*

[234] Allison Lewis, *The NYPD's New DNA Dragnet: The Department is Collecting and Storing Genetic Information, With Virtually No Rules to Curb Their Use*, N.Y. DAILY NEWS (Feb. 8, 2019, 6:52 PM), https://www.nydailynews.com/opinion/ny-oped-the-nypds-new-dna-dragnet-20190206-story.htm [https://perma.cc/5SQY-SCNF].

[235] The case, *Lowe v. Atlas Logistics Group Retail Services*, 102 F. Supp. 3d 1360 (N.D. Ga. 2015), soon became colloquially known by this name in the popular press.

[236] Defendant's Motion for Summary Judgment at 2, *Lowe*, 102 F. Supp. 3d at 1360.

[237] *Lowe*, 102 F. Supp. 3d at 1361.

GINA is meant to address the risks of health insurers and employers discriminating against people on the basis of genetic tests[238] and family medical history.[239] Privacy is conceptualized in the law not as an end in itself,[240] but as a "bulwark [to prevent] access to the very information health insurers or employers could use to discriminate."[241] The core concern to which the law is addressed is the worry that health insurers might raise premiums or drop coverage based on health risks revealed by genetic data; the law's application to employers is premised on the connection between employment and health insurance cost.[242] GINA addresses these risks by preventing health insurers from making eligibility or premium determinations on the basis of genetic information;[243] by making it unlawful for employers to discriminate on the basis of genetic information; and by making it unlawful for an employer to request or require information from an employee or an employee's family member.

In other words, GINA was initially devised as an intervention on two types of dependency. It addresses *similarity-based* dependencies because genetic information may allow for inference based on the predispositions and health outcomes of known individuals with comparable profiles. And it speaks to *tie-based* dependencies because information about the genetic profiles and manifested health conditions of family members may be used to learn more about the employee—for example, the risk of hereditary diseases. In both cases, the goal of the statute is to guard against discrimination on the basis of genetic conditions or predispositions.

Yet in the devious defecator case, the plaintiffs made a different argument about GINA's protections. In their case, the dependency at stake was not based on similarity or tie but was a *difference-based* dependency. By drawing up a list of suspected defecators based on work schedules—and then winnowing the list by asking employees to provide DNA samples to exclude themselves from suspicion—the company sought to use employees' DNA for identification, rather than for discrimination

[238] By its terms, GINA offers protection for a person's own genetic tests and the genetic tests of her family members. 29 U.S.C. § 1182 (2018).

[239] In addition to a family member's genetic tests, GINA offers protection against discrimination on the basis of the manifested health conditions of a person's family members, including dependents and relatives up to four degrees away. *Id.* GINA does *not* provide a cause of action for disparate impact discrimination based on genetic information, in contrast of other civil rights statutes. *See* Ifeoma Ajunwa, *Genetic Data and Civil Rights*, 51 HARV. C.R.-C.L. L. REV. 75 (2016) (arguing for the authorization of a disparate impact cause of action).

[240] Bradeley A. Areheart & Jessica L. Roberts, *GINA, Big Data, and the Future of Employee Privacy*, 128 YALE L.J. 711, 715 (2019).

[241] *Id.* at 718.

[242] *Id.* at 723–24.

[243] Note that this component of GINA is an example of a "don't use" restriction, as discussed in Peppet, *supra* note 174, at 1200 *et seq*.

based on the medical information it contained. The crux of the case came down to the court's analysis of whether GINA should be interpreted to protect workers against this sort of use. The warehouse company argued that the statute should be interpreted *only* to bar employers from using information related to an individual's propensity for disease.[244] Yet the plain meaning of the statute, as the employees argued, prohibits employers from "request[ing], requir[ing], or purchas[ing] genetic information with respect to any individual prior to such individual's enrollment under the plan or coverage in connection with such enrollment,"[245] save for certain inapplicable exceptions. The jury agreed with this broader interpretation and awarded the employees $2.2 million in damages.[246]

Our taxonomy of the relevant privacy dependencies explains why many were surprised by the case's outcome. Despite the court's statutory construction legitimating the plaintiffs' claims, it seems quite clear that GINA was conceptualized as a route to protect against similarity- and tie-based dependency. In response to the ruling, biotechnology law expert John Conley opined that "[t]his is an application of the law that no one thought of in a million years . . . . But the ruling is not controversial. You can't use genetic testing for dismissal purposes."[247]

The court's application of GINA to a *difference-based* dependency does more than nominally augment the statute's purview; it suggests that the law might protect a wholly different and more expansive set of normative values than those initially conceived. Though the law on its face (and by its title) is addressed to concerns about discriminatory treatment in rate-setting and to prediction based on genetic data[248]—both forward-looking uses involving actuarial assessment about what an employee will do (and cost) in the future—the devious defecator case has nothing to do with discrimination or prediction. Instead, it suggests that employers may not use employees' DNA for forensic investigation about *past* events, or for the purposes of identifying and disciplining employees for such ostensible misbehavior. This use transforms GINA from a

[244] *Lowe*, 102 F. Supp. 3d at 1365.

[245] 29 U.S.C. § 1182 (2018).

[246] Areheart & Roberts, *supra* note 240, at 752.

[247] Ajunwa, *supra* note 239, at 113 (quoting Natasha Gilbert, *Why the 'Devious Defecator' Case is a Landmark for US Genetic-Privacy Law*, NATURE (June 25, 2015), http://www.nature.com/news/why-the-devious-defecator-case-is-a-landmark-for-us-genetic-privacy-law-1.17857 [https://perma.cc/42MZ-2DP2]).

[248] It is worth noting that GINA's success as a nondiscrimination statute is equivocal. Areheart and Roberts's empirical survey of cases brought under GINA demonstrated that *no* successful GINA claims have been premised on discrimination based on the results of genetic tests. Areheart & Roberts, *supra* note 240, at 714.

relatively narrow risk allocation statute to a "robust protection for employee privacy."[249]

IV. CONCLUSION

Unpacking the disparate mechanisms that create different privacy dependencies can give us the necessary clarity for policymaking and regulation. Most immediately, it can help us determine if and when we even recognize Bob as a party with a legitimate privacy claim when Alice is the disclosing or observed party. It can shed light on the varied normative goals that we expect privacy to serve under different configurations. Finally, attending to the specific relationships that create privacy dependencies can suggest possible targets for intervention—opportunities to capitalize on mutual dependency or ensure greater independence.

Certain forms of dependency afford greater opportunity for social solidarities to develop than others do. If people are made aware of how their disclosures may implicate close social ties, they may refrain from making such disclosures. Inference on the basis of socially salient characteristics, particularly those protected by discrimination law, might be countered by activism and advocacy. People wishing to preserve anonymity within a group might rely on collective action to make themselves less readily distinguishable. But other dependencies make privacy-protective solidarity less likely, like inference on the basis of non-socially-salient characteristics and disclosures that involve distinguishing oneself from a group for favorable treatment.

Privacy dependencies should thus not only call into question notice and choice as a model for privacy regulation; they should force us to abandon the naïve hope that solidarity can help rescue informed consent by clarifying the degree to which our privacy choices implicate others. If we are scarcely able to make decisions that attend to our own privacy interests, the goal of recognizing shared interests should not be to further burden our individual choices with an expectation that we take into account the interests of others.[250] At its best, solidarity can foster collective action demanding technologies, policies, and laws that address the mechanisms that create dependencies, relieving individuals of the

[249] Areheart & Roberts, *supra* note 240, at 752. Areheart and Roberts go on to note that the devious defecator case clarifies the "independent moral value" of workplace privacy as an intrinsic harm, analytically separable from the extrinsic harm of potential discrimination based on private information. *Id.* at 779–80.

[250] *Contra* Carissa Véliz, *Privacy is a Collective Concern*, NEW STATESMAN (Oct. 22, 2019), https://www.newstatesman.com/science-tech/privacy/2019/10/privacy-collective-concern [https://perma.cc/PGS2-FZJS] (arguing that dependencies make people morally responsible for one another's privacy).

impossible task of managing collective interests through their individual decisions. Recognizing the mechanisms that create different forms of dependency does more than demonstrate the shortcomings of privacy individualism; it lays the groundwork for well-tailored policymaking and advocacy.