TRANSPARENT PREDICTIONS

*Tal Z. Zarsky**

Can human behavior be predicted? A broad variety of governmental initiatives are using computerized processes to try. Vast datasets of personal information enhance the ability to engage in these ventures and the appetite to push them forward. Governments have a distinct interest in automated individualized predictions to foresee unlawful actions. Novel technological tools, especially data-mining applications, are making governmental predictions possible. The growing use of predictive practices is generating serious concerns regarding the lack of transparency. Although echoed across the policy, legal, and academic debate, the nature of transparency, in this context, is unclear. Transparency flows from different, even competing, rationales, as well as very different legal and philosophical backgrounds. This Article sets forth a unique and comprehensive conceptual framework for understanding the role transparency must play as a regulatory concept in the crucial and innovative realm of automated predictive modeling.

Part II begins by briefly describing the predictive modeling process while focusing on initiatives carried out in the context of federal income tax collection and law enforcement. It then draws out the process's fundamental elements, while distinguishing between the role of technology and humans. Recognizing these elements is crucial for understanding the importance and challenges of transparency. Part III moves to address the flow of information the prediction process generates. In doing so, it addresses various strategies to achieve transparency in this process—some addressed by law, while others are ignored. In doing so, the Article introduces a helpful taxonomy that will be relied upon throughout the analysis. It also

* Associate Professor, University of Haifa, Faculty of Law. Hauser Research Fellow, New York University Law School, 2010–11. Columbia Law School, LLM (2001), JSD (2004). I thank Danielle Citron, Mark Fenster, James Grimmelmann, Chaim Saiman, Helen Nissenbaum, Frank Pasqualle, Paul Rosenzweig, Kathy Strandburg, Erin Murphy, Ira Rubinstein, and the members of the Privacy Reading Group at NYU Law School and the Villanova Law School Faculty Seminar for their comments. Research for this paper was partially funded by an NWO (the Dutch Research Foundation) grant titled: "Data Mining without Discrimination," and I thank my co-researchers Bart Custers, Bart Schermer, and Toon Calders for their insights. I also thank Ran Wolff for important insights regarding the technical aspects of this project and Talya Ponchek for her assistance with this Article.

establishes the need for an overall theoretical analysis and policy blueprint for transparency in prediction.

Part IV shifts to a theoretical analysis seeking the sources of calls for transparency. Here, the analysis addresses transparency as a tool to enhance government efficiency, facilitate crowdsourcing, and promote both privacy and autonomy. Part V turns to examine counterarguments which call for limiting transparency. It explains how disclosure can undermine government policy and authority, as well as generate problematic stereotypes. After mapping out the justifications and counterclaims, Part VI moves to provide an innovative and unique policy framework for achieving transparency. It concludes, in Part VII, by explaining which concerns and risks of the predictive modeling process transparency cannot mitigate, and calling for other regulatory responses.

TABLE OF CONTENTS

I. INTRODUCTION: TRANSPARENCY AND THE "DATA EYE IN THE SKY"

Can machines predict human behavior? A broad variety of governmental initiatives are setting out to try.[1] Recent advances in mathematics, artificial intelligence, and computer science might bring society closer to achieving this futuristic objective.[2] Vast datasets of personal information available to commercial and governmental entities enhance the government's ability to engage in these ventures and the appetite to push them forward.[3]

A recent *New York Times* report provides a glimpse of what the future might hold, describing a U.S. intelligence proposal to build a "Data Eye in the Sky."[4] This initiative will structure vast databases of information collected from a broad variety of digital sources, such as Internet traffic, web searches, and Twitter and Facebook posts.[5] Thereafter, it will continuously analyze this data, striving to identify various trends. These trends will later be used to predict the spread of pandemics as well as future political and economic developments.[6]

We need not look into the future for automated predictive schemes. Several are already used today. Law enforcement and security forces are developing, implementing, and already utilizing individualized automated prediction tools.[7] Among others, they use personal information collected regarding specific individuals to make educated guesses as to the individuals' next steps and the risks they might hold.[8] Indeed, governments have a distinct interest in automated individualized predictions to foresee unlawful actions. This is especially true when the unlawful behavior can cause considerable damage or is difficult to police. There is no doubt that these predictive measures will continue to penetrate other aspects of life that feature interactions between government and the public.[9]

[1] *See infra* Part II.

[2] *See* John Markoff, *Government Aims to Build a 'Data Eye in the Sky,'* N.Y. TIMES, Oct. 10, 2011, at D1.

[3] *Id.*

[4] *Id.*

[5] *Id.*

[6] *Id.*

[7] Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. & HIGH TECH. L. 235, 235–36 (2011).

[8] *Id.*

[9] For a similar view, see *id.* at 237–39.

Predictive analysis schemes quickly bring the *Minority Report* dystopia to mind,[10] in which individuals are sanctioned and imprisoned for crimes they have yet to commit. Governments today are far less aggressive in their responses to the outcomes of predictive analysis. Yet, they are increasingly curious to figure out what we will do next and take action, rather than wait to investigate what has already happened and suffer the possible consequences.

The growing use of predictive practices, premised upon the analysis of personal information and powered by data mining, has generated a flurry of negative reactions and responses.[11] Individuals, watchdog groups, and policy makers all fear that the process might impede autonomy and privacy, will be biased and discriminatory, and could be tainted with errors and overinvasive.[12] They also fear it might quickly spread from limited, approved realms to many other tasks and arms of governments.[13] Yet an overall concern is the lack of transparency these processes entail.[14] A call for transparency emerges from the public, from the press,[15] and even from the legislator.[16] The call for transparency is commonly cited when requiring changes in these initiatives.[17] To a great extent, the lack of transparency feeds, or even initiates, many of the other concerns mentioned.

Although echoed across the policy, legal, and academic debate, the nature of transparency in the context of governmental automated predictions is unclear and calls for a rigorous analysis. In the context of

[10] PHILIP K. DICK, THE MINORITY REPORT (2002).

[11] *See* Anita Ramasastry, *Lost in Translation? Data Mining, National Security and the "Adverse Inference" Problem*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 757, 760 (2004) (discussing the reaction to the "Total Information Awareness" initiative); Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317 (2008); *see also* CRS REPORT, *infra* note 63, at 5–11.

[12] For an account of these analyses, see Tal Z. Zarsky, *"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1, 18–50 (2002); *see also* TECHNOLOGY AND PRIVACY ADVISORY COMMITTEE, SAFEGUARDING PRIVACY IN THE FIGHT AGAINST TERRORISM (2004) [hereinafter TAPAC REPORT], *available at* http://epic.org/privacy/profiling/tia/tapac_report.pdf; JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 167–69, 256 (2012); Daniel J. Solove, *Data Mining and the Security-Liberty Debate*, 74 U. CHI. L. REV. 343 (2008); James X. Dempsey & Paul Rosenzweig, *Technologies That Can Protect Privacy as Information Is Shared to Combat Terrorism*, LEGAL MEMORANDUM, No. 11, May 26, 2004, *available at* James X Dempsey link: http://www.heritage.org/research/reports/2004/05/technologies-that-can-protect-privacy-as-information-is-shared-to-combat-terrorism.

[13] This is the notion of mission creep. *See* Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435, 485 (2008); Slobogin, *supra* note 11, at 326.

[14] *See* Pasquale, *supra* note 7, at 237.

[15] Todd Essig, *'Big Data' Got You Creeped Out? Transparency Can Help*, FORBES, Feb. 27, 2012, http://www.forbes.com/sites/toddessig/2012/02/27/big-data-got-you-creeped-out-transparency-can-help/.

[16] Federal Agency Data Mining Reporting Act of 2007, 42 U.S.C. § 2000ee-3(c)(2) (Supp. III 2007).

[17] TAPAC REPORT, *supra* note 12, at 49–50; COHEN, *supra* note 12, at 234–39; Cate, *supra* note 13, at 481; Solove, *supra* note 12, at 359–61.

automated prediction, transparency flows from different, even competing, rationales, as well as a variety of legal and philosophical backgrounds. Theories related to government fairness and efficiency, innovation policy, privacy, autonomy, and the problems of stigma and stereotyping all must be accounted for. When viewed in concert, they lead to different, and at times contradicting, conclusions and practical recommendations. This Article sets out to establish a coherent understanding of what transparency both means and entails in this crucial context, while relying on the legal literature set forth in all these fields. To understand transparency in this unique context, the Article sets forth a conceptual framework, drawing from the growing field of transparency-related legal scholarship and providing an innovative theoretical taxonomy. It works through the various stages of automated prediction and addresses the different transparency rationales relevant to every juncture.

The notion of transparency has recently been attracting attention in legal scholarship.[18] Transparency or related terms such as "Freedom of Information"[19] and "Open Government"[20] were hailed by President Obama and others as cornerstones of the current administration.[21] More specifically, transparency is already addressed by law in the context of predictive analysis. The Privacy Act, the Freedom of Information Act (FOIA), the Federal Agency Data Mining Reporting Act, and the E-Government Act all mandate transparency, while addressing various facets of the automated prediction process.[22] Additional agency-specific regulations further promote transparency.[23] Yet even with this extensive legal mechanism in place, the current framework does not provide a sufficient, balanced, and nuanced response to today's challenges. On the

[18] *See, e.g.*, Frederick Schauer, *Transparency in Three Dimensions*, 2011 U. ILL. L. REV. 1340 n.10 (providing an extensive list of recent publications). In addition, recently, on his popular "Legal Theory Blog" Larry Solum chose to highlight this concept. Lawrence Solum, *Legal Theory Lexicon: Transparency*, LEGAL THEORY BLOG (Jan. 22, 2012, 5:53 PM), http://lsolum.typepad.com/legaltheory/2012/01/legal-theory-lexicon-transparency.html.

[19] Freedom of Information Act, 5 U.S.C. § 552 (2006).

[20] *See, e.g.*, *Freedom of Information Act Memorandum for the Heads of Executive Departments and Agencies*, 74 Fed. Reg. 4683 (Jan. 21, 2009), *available at* http://edocket.access.gpo.gov/2009/pdf/E9-1773.pdf.

[21] This concept of transparency is also advocated in the context of large private entities. These discussions will not be the focus of this Article, yet I will at times refer to the rationales and arguments stated in the broader discussion of transparency and examine how they can be applied to the governmental context. *See, e.g.*, LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT (1914); Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669 (2010).

[22] *See* The Privacy Act of 1974, 5 U.S.C. § 552a (2006); 5 U.S.C. § 552; 42 U.S.C. § 2000ee-3; E-Government Act of 2002, 44 U.S.C. § 3501, 3601 (2006).

[23] The Department of Homeland Security (DHS) issued privacy guidelines and a privacy policy by which it must abide. Memorandum from Hugo Tevfel III, Chief Privacy Officer, U.S. Department of Homeland Security, on The Fair Information Practice Principles: Framework for Privacy Policy at the Department of Homeland Security (Dec. 29, 2008) [hereinafter DHS GUIDELINES], *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf.

basis of an innovative theoretical model hereby developed, this Article strives to draw out a policy blueprint addressing the notion of transparency as it emerges in the predictive modeling context.

Closely examining transparency in predictive modeling is merely a subset of the growing study of transparency and technological change.[24] A discussion of this theme balances great promises against grim threats. Communication-based and search-enabling technologies generate the promise of transparency by efficiently conveying information to citizens[25] and bringing them into the public discourse.[26] Achieving transparency, however, is easier said than done.[27] The process of merely flooding the public with facts and figures does not effectively promote transparency. It might even backfire.[28] The challenge of achieving transparency in the digital era led to a worldwide effort to establish suitable legal and technological formats, and this context has already generated an acronym—"TETs" (Transparency Enhancing Tools)—for applications striving to meet this objective.[29] Yet technological and institutional design must first rely upon a solid philosophical and policy discussion, which this Article sets out to address.

Technology and transparency have yet another, darker, connection. Governments and other large institutions employ computers for support, advice, and even decision making. They are learning to rely on these systems, at times almost blindly.[30] The automated technological process and outcome might be considered as flawless and neutral, yet commentators are pointing out that they are neither.[31] Furthermore, scholars point out that that they are also opaque by nature. Scholarship in the context of risk analysis,[32] electronic voting machines,[33] search

[24] *See* ARCHON FUNG ET AL., FULL DISCLOSURE: THE PERILS AND PROMISE OF TRANSPARENCY 12–15 (2007).

[25] *Id.*; *see* 44 U.S.C. § 3501; *Office of E-Government & Information Technology*, THE WHITE HOUSE, http://www.whitehouse.gov/omb/e-gov/ (last visited Feb. 22, 2013). For advances in the mobile realm, see José M. Alonso et al., *Improving Access to Government Through Better Use of the Web*, W3C (May 12, 2009), http://www.w3.org/TR/2009/NOTE-egov-improving-20090512/.

[26] This is a process also referred to as E-Gov 2.0. *See* Viktor Mayer-Schönberger & David Lazer, *E-Gov and the Coming Revolution of Information Government* (Belfer Center Sci. Int'l Affairs, Working Paper), *available at* http://live.belfercenter.org/files/intro-wp.pdf; Richard MacManus, *E-Government Meets Web 2.0: Goodbye Portals, Hello Web Services*, READWRITE (Nov. 5, 2007), http://www.readwriteweb.com/archives/e-government_meets_web_20.php.

[27] In the context of participation in rulemaking, see Beth Simone Noveck, *The Electronic Revolution in Rulemaking*, 53 EMORY L.J. 433, 436–37 (2004). *See generally* Viktor Mayer-Schönberger, *Information Government and the Locus of Implementation* (Belfer Center Sci. Int'l Affairs, Working Paper), *available at* http://belfercenter.ksg.harvard.edu/files/vms.pdf.

[28] Jennifer Shkabatur, *Cities @ Crossroads: Digital Technology and Local Democracy in America*, 76 BROOK. L. REV. 1413, 1437 (2011).

[29] Mireille Hildebrant, *Profiling and the Rule of Law*, 1 IDENTITY INFO. SOC'Y 55, 66 (2008).

[30] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271 (2008) [hereinafter Citron, *Technological Due Process*].

[31] Bamberger, *supra* note 21, at 675; Citron, *Technological Due Process*, *supra* note 30, at 1256–58.

[32] Bamberger, *supra* note 21, at 739.

engines,[34] and automated administrative decisions[35] all point to the troubles of automated decision making and how they are further complicated by the opacity of technology. Yet referring to such opacity in general is insufficient. This Article draws out a detailed taxonomy to articulate the implications of such opacity in the specific context of automated predictions, and how it could be resolved.

Part II begins by briefly describing and explaining the practices of predictively modeling individual conduct. It does so by focusing on practical examples—federal income tax collection and law enforcement.[36] It then generally draws out the specific elements of the automated prediction process while distinguishing between the role of technology and humans. Part III strives to achieve two important objectives. It moves to address the flow of information generated in the prediction process. Along the way, it sets forth the forms of transparency addressed by law—as well as other ways transparency could be understood and achieved—which the law currently overlooks. In doing so, this Part demonstrates the need for an overall theoretical analysis and policy blueprint and introduces a helpful taxonomy regarding the various segments of the prediction process.

Part IV provides a theoretical overview, seeking the unique foundations for transparency in automated predictive modeling. Transparency is grounded in theories explaining it as an important measure to enhance government efficiency, to facilitate crowdsourcing, to protect the privacy of relevant data subjects, and to enhance the autonomy of those affected by the process. Part V turns to examine transparency's negative, and at times unintended, consequences in this context. It explains how disclosure can undermine both government policy and authority, as well as generate stereotypes. Both Parts IV and V also further examine the implication and strengths of these theories at every segment of the predictive process, while drawing out crucial distinctions relevant to this unique context, which have yet to be addressed in the literature.

After mapping out the justifications and counterclaims, Part V moves to provide an innovative and unique policy blueprint for achieving transparency in prediction modeling. It does so by providing concrete recommendations, premised on balancing the previously discussed theories. Only after mapping out the theoretical elements can we establish what the nature of the call for transparency is and where it must lead us. Further, it is only after working through this maze of arguments that we can understand what risks transparency does *not* mitigate, and

[33] Danielle Keats Citron, *Open Code Governance*, 2008 U. CHI. LEGAL F. 355, 357 [hereinafter Citron, *Open Code Governance*].

[34] *See* Pasquale, *supra* note 7, at 244.

[35] Citron, *Technological Due Process*, *supra* note 30, at 1279.

[36] *See infra* Part II. A–B.

which still remain lingering and requiring separate responses—a point the Article concludes with in Part VII.

II. PREDICTIONS IN THE AUTOMATED ADMINISTRATIVE STATE

The use of automated prediction is an intriguing dynamic which carries with it the promises of efficiency and overall success.[37] It holds the potential of spreading throughout the administrative state. These predictions call for the use of sophisticated computer programs and extensive datasets, as well as professional data analysts.[38] The process relies upon various assumptions pertaining to society and human nature. To understand these points, this analysis begins by examining two notable governmental predictive tasks. Thereafter, this Part moves to take a broader view of the prediction dynamic, while studying the *technology* enabling these projects, the role of the *human* analyst in what seems to be an automated process, and the *policy decisions* underlying many of the steps of these processes.

A. The Internal Revenue Service and Selective Auditing Strategies

We begin with a simple example, yet a painful topic for almost all—taxes. It is no secret that almost nobody likes paying them, and many evade their payment using various strategies. The Internal Revenue Service (IRS) receives millions of tax returns annually. Given their limited manpower, IRS agents manually audit a mere fraction of the forms received.[39] The IRS tries to deter tax evaders by inflicting harsh penalties on those it catches.[40] Yet, the very low chance of being singled out by the IRS as well as the inability to punish non-proportionately still leads many individuals to cheat.[41] Thus, the IRS turns to prediction models.[42] Given the vast amount of information at its disposal, the IRS relies on computerized automation for assistance in the prediction process (while supplementing this process with random auditing).[43]

[37] Bamberger, *supra* note 21, at 685–87.

[38] *Id.*

[39] The accepted estimate is one percent for individuals. These facts are frequently published by the IRS. For 2007, see *Fiscal Year 2007 Enforcement and Services Results*, INTERNAL REVENUE SERVICE, http://www.irs.gov/uac/Fiscal-Year-2007-Enforcement-and-Services-Results (last updated Sept. 6, 2013) [hereinafter 2007 Tax Statistics].

[40] FREDERICK SCHAUER, PROFILES, PROBABILITIES AND STEREOTYPING 161 (2003).

[41] *Id.* at 161–62.

[42] I acknowledge that the IRS example is somewhat problematic; the IRS auditing process is not only intended to detect cheating, but also to generate deterrence. For that, the IRS engages in a variety of strategies which includes the manipulative use of the media. For this discussion, I am setting aside this aspect of the IRS enforcement actions, and focusing on the mere struggle to detect cheating to the greatest extent possible. For more on the IRS's latter strategies, see Joshua D. Blank, *In Defense of Individual Tax Privacy*, 61 EMORY L.J. 265, 318 (2011).

[43] For instance, to sustain a dataset so as to later find trends for auditing or to examine whether compliance levels and patterns changed—for an argument that only random examinations of returns

The IRS uses predictive modeling to find those most likely to file false returns. It does so while analyzing the vast datasets at its disposal, which include information regarding previous instances of tax fraud, and other personal and financial data.[44] The IRS uses an automated process to construct a dynamic pattern or profile of individuals who have a higher chance of evading taxes.[45] Those indicated in the process are subjected to additional scrutiny and auditing.[46] Selection for manual auditing has severe implications. The auditing process is known to be unpleasant, stressful, time consuming, and costly.[47] Individuals selected by the prediction model will *not*, however, automatically be considered as suspects or criminals.[48] It is also doubtful whether a negative stigma attaches to the audited individual.[49] The IRS is not alone in carrying out such predictions. Other agencies charged with allocating government benefits are following suit and carrying out similar analyses.[50]

The IRS prediction process is not flawless. It generates false positives—which refer to those who are part of the "scrutinized" pattern, yet filed truthful returns.[51] It also includes false negatives—those outside the pattern, who nevertheless chose to cheat on their returns. I will note the existence of these forms of errors throughout my analysis.

would achieve a fair and effective outcome, see BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 238 (2007).

[44] For more on this dynamic, see SCHAUER, *supra* note 40, at 163–67.

[45] As the IRS explains: "We accept most taxpayers' returns as filed. If we inquire about your return or select it for examination, it does not suggest that you are dishonest. The inquiry or examination may or may not result in more tax. We may close your case without change or you may receive a refund. The process of selecting a return for examination usually begins in one of two ways. One way is to use computer programs to identify returns that may have incorrect amounts. The programs may be based on information returns, such as Forms 1099 or W-2, on studies of past examinations, or on certain issues identified by other special projects. Another way is to use information from compliance projects that indicates a return may have incorrect amounts. These sources may include newspapers, public records, and individuals. If we determine the information is accurate and reliable, we may use it to select a return for examination." I.R.S. Publication 3498 (Rev. 11-2004), *available at* http://www.irs.gov/pub/irs-pdf/p3498.pdf; *see also* I.R.S. Publication 1 (Rev. Sept. 2012) [hereinafter Taxpayer Rights], *available at* http://www.irs.gov/pub/irs-pdf/p1.pdf. The system is commonly referred to as the *DIF score*.

[46] Some refer to the elements used as "audit triggers." *See* Andrea Coombes, *IRS Audit Triggers and Red Flags—the 2010 Tax Guide from MarketWatch*, MARKETWATCH (Apr. 13, 2009, 10:27 AM), http://www.marketwatch.com/story/want-avoid-audit-consider-how; *Tax Audit Triggers to Watch Out For*, BOSTON.COM, http://www.boston.com/business/taxes/articles/macpa/new_2005/Tax_audit_triggers/ (last visited Feb. 22, 2013) [hereinafter *Tax Audit Triggers*].

[47] *See Tax Audit Triggers*, *supra* note 46.

[48] *See supra* note 45.

[49] *See* Blank, *supra* note 42, at 286.

[50] *See* U.S. GENERAL ACCOUNTING OFFICE, GAO-04-548 DATA MINING: FEDERAL EFFORTS COVER A WIDE RANGE OF USES (2004) [hereinafter GAO REPORT], available at http://www.gao.gov/new.items/d04548.pdf. For instance, veteran data is mined to search for frauds and abuses. Additional projects for the detection of fraud are listed. There is a long list of planned projects that involve data mining and are destined to increase tax compliance. *See id.* at 52; *see also* Citron, *Technological Due Process*, *supra* note 30, at 1263.

[51] *See* Cate, *supra* note 13, at 475 (discussing the inevitability of false positives and their impact).

The IRS example seems to be an awkward choice to commence a discussion on transparency in prediction—especially because of the distinct *lack* of transparency throughout this process. The IRS maintains full secrecy as to the selective auditing schemes it applies.[52] While it does publish its overall auditing statistics,[53] it does not reveal the methods it applies to predict audited returns or the actual factors it uses. The public cannot establish the false positives and negatives in this process.

Yet this example is extremely interesting for this precise reason. The IRS example demonstrates how elusive transparency can be in the predictive realm. The process here described is not esoteric. It pertains to almost all Americans. Indeed, the opacity of this process has not gone unnoticed. A specific 1998 legislative amendment was meant to generate transparency.[54] Yet, as we will now see, because of poor planning and perhaps a lack of theoretical foundations, it was easily circumvented.

A specific section of the IRS Restructuring and Reform Act, titled "Disclosure of Criteria for Examination Selection,"[55] calls for the publication of general criteria and informing the audited individual of the factors that triggered the audit. The law states, however, that both requirements must be met without impeding law enforcement.[56]

In practice, the implementation of transparency according to this law is laughable. The first legislative requirement ("general criteria") is met by statements added to the "Declaration of Taxpayer's Rights," explaining very generally how examinations are selected.[57] Reviewing

[52] SCHAUER, *supra* note 40, at 163.

[53] *See* 2007 Tax Statistics, *supra* note 39.

[54] Several cases address a closely related topic—exposing an individual's DIF score. Plaintiffs often challenge the IRS with FOIA requests to reveal their DIF score. Courts have systematically denied such requests, while noting several exemptions, most notably 26 U.S.C. § 6103(b)(2), 5 U.S.C. § 552(b)(3) (both of which rely upon instances where a specific statute allows withholding information—a statute the court recognizes in these cases), and 5 U.S.C. § 552(b)(7)(the "law enforcement exemption"). *See* Huene v. U.S. Dep't of the Treasury, No. 2:11–cv–02109 JAM KJN PS, 2012 WL 3730635, at *6 (E.D. Cal. Aug. 24, 2012). "Courts have concluded that the release of a taxpayer's DIF scores could reasonably be expected to risk circumvention of the law, as provided in 5 U.S.C. § 552(b)(7)(E), in that the release of such scores could enable taxpayers to determine how to lower DIF scores in order to avoid audits." *Id.* at *7. Courts have also recognizes the IRS's right to withhold other sources of information so not to allow others to understand the nature of the DIF formula. *Id.* More generally, current case law allows government to deny FOIA requests when these might undermine law enforcement attempts (and these denials were applied in several instances, such as in the process of detecting Medicare fraud). *See, e.g.*, Gerstein v. CIA, No. C-06-4643, 2010 WL 669743 at *4 (N.D. Cal. Feb. 23, 2010) ("In *Dirksen v. United States Department of Health and Human Services* . . . for example, the agency was found to have properly applied Exemption 2 to withhold the guidelines used to categorize, for speed in processing, the types of Medicare claims that could be granted automatically as well as those that either would be denied or subject to more detailed review . . . . The court in *Dirksen* reasoned that such withholding was proper under Exemption 2 because public disclosure would allow individuals to fashion their claims to conform to the guidelines, thereby causing the guidelines to 'lose the utility they were intended to provide.' (citations omitted)").

[55] Internal Revenue Service Restructuring and Reform Act of 1998, Pub.L. No. 105–206, 112 Stat. 685, 771 (1998) (codified as amended at 26 U.S.C. 7801 (2006)).

[56] *Id.*

[57] Taxpayer Rights, *supra* note 45.

these clauses reveals close to nothing. The second legislative requirement, "informing the audited individual," is met by providing the audited individual with information as to whether the audit was selected at random, chosen by the computer software, or as a result of external sources, such as the media.[58] Again, such limited disclosure is useless.

There is another side to the transparency/opacity-in-auditing story worth noting. Based on anecdotal information, accountants are able to construct what they believe to be "audit triggers."[59] Accountants carefully note thresholds for itemized deductions and expenses that they deem "too high" or reported salaries and income that are too low. They can arguably point to professions that are more likely to be audited, such as those operating in cash-based business (waiters, contractors and barbers are famous examples) as well as others (lawyers and physicians).[60] They identify deductions that generate suspicion (such as the one for a home office).[61] Therefore, alternative flows of information (and their elaborate effects on the distribution of knowledge and wealth) are forever lurking in the background, even when a process is not officially rendered transparent.

The IRS example teaches us three central lessons. First, that automated predictions are upon us, affecting our lives constantly even without us noticing. Second, that the lack of transparency in these practices generates serious concerns that mobilize the public and lead to regulatory and even legislative responses. Third, that without proper understanding of how the process works and what transparency should be achieving, opacity will remain intact.

B. Data Mining and Security

Since the events on September 11, 2001 and subsequent attacks around the world, governments are working extremely hard to preempt

[58] *See* IRS, INTERNAL REVENUE MANUAL 10.8.19.1 (Aug. 11, 2006), *available at* http://www.irs.gov/irm/part4/irm_04-010-008-cont02.html. ("The taxpayer may inquire about why his/her return was selected for examination. Publication 1, *Your Rights as a Taxpayer*, has been revised and includes a statement describing the criteria and general procedures for selecting taxpayers for examination. The Service is not required to disclose the basis for the selection of a particular taxpayer for examination. Generally, it is the practice of the Service to respond if the source of the examination is random, DIF generated (without explaining the scoring process), or if generated from a public source (e.g., public media report). However, if the source of the examination is an informant, the Service is not obligated to, nor would it be appropriate to, disclose an informant exists. The examiner and his manager should consult with Disclosure when requested to provide a response to return selection for informant cases.").

[59] Christopher A. Szechenyi, *Talking Taxes-How to Avoid an Audit*, SALARY.COM, http://www.salary.com/Articles/ArticleDetail.asp?part=par258 (last visited May 20, 2013); *see also Tax Audit Triggers supra* note 46; SCHAUER, *supra* note 40, at 165 (listing additional references and examples).

[60] *See supra* note 59.

[61] *Id.*

future attacks.[62] Among various initiatives, it is reported that governments are employing predictive data mining to study trends in the actions of attackers and attacks.[63] With predictive models in hand, individuals identified as higher risks are contacted or set aside for further questioning or scrutiny.

The public is learning of these practices directly from the government[64] or when they are leaked to the press.[65] Yet there is still an overall sense that much about these projects remains secret. In an infamous incident, the public reacted with awe to the Total (and later "Terrorism") Information Awareness (TIA) project.[66] Parts of this project presumably called for predictive data mining premised upon both public and personal information.[67] It is fair to assume that lack of transparency in these projects contributed to the public's negative response. This project was famously halted by Congress.[68] The development of similar projects continues, however, under other names and acronyms.[69] Beyond TIA, more limited ventures using similar techniques were set in place. The most famous and salient examples pertain to *airports*.

At airports, the interaction between individuals and security forces takes place in at least two contexts: securing airline travel and stopping unwanted individuals from entering (or in some cases, exiting) the country. Initiatives pertaining to the first context use the analysis of personal information to identify risk. By examining information available from previous attacks, the government considered constructing prediction models and profiles to identify specific "risk-related" individuals.[70] Indeed, the CAPPS II project was intended to do so while relying upon governmental and possibly even commercial databases.[71] This project was cancelled, among others, in view of privacy and transparency concerns. The government followed up this attempt by

[62] *See* JEFFREY W. SEIFERT, CONG. RESEARCH SERV., RL31798, DATA MINING AND HOMELAND SECURITY: AN OVERVIEW 8 (2008) [hereinafter CRS REPORT], *available at* http://www.fas.org/sgp/crs/homesec/RL31798.pdf.

[63] Homeland Security Act of 2002, Pub.L. No. 107-296, 116 Stat. 2135, 2147 (2002), (authorizing DHS to make use of data mining to achieve its objectives); s*ee* 6 U.S.C. § 121(d) (13) (2006).

[64] U.S. DEP'T. OF HOMELAND SECURITY PRIVACY OFFICE, 2009 DATA MINING REPORT TO CONGRESS 2 (2009), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_datamining_2009_12.pdf.

[65] Famous examples are the National Security Agency (NSA) "blanket" wiretapping initiatives, which generated subsequent lawsuits against telecom operators. *See* Cate, *supra* note 13, at 448.

[66] *Id.* at 449.

[67] *Id*.

[68] *Id.* at 451.

[69] Laura K. Donohue, *Criminal Law: Anglo-American Privacy and Surveillance*, 96 J. CRIM. L. & CRIMINOLOGY 1059, 1145 (2006); Slobogin, *supra* note 11, at 318.

[70] Donohue, *supra* note 69, at 1136.

[71] Slobogin, *supra* note 11, at 322.

developing the "Secure Flight" protocol.[72] This project was stalled several times because of privacy concerns but is now probably moving ahead.[73] In its current stage, this project does not involve predictive modeling. It merely matches personal information travelers provide to a "No Fly List"[74] prepared by the authorities.[75] It is unclear whether predictive data mining is used in the preparation of the "No Fly List." One should not be surprised, however, if automated prediction processes are applied in this context, given the governmental appetite for it.

Automated prediction is used in airports to achieve law enforcement's second objective—protecting borders. The Department of Homeland Security (DHS) is currently using (and further developing) predictive modeling, which is powered by data mining to secure the exit and entry of individuals into and from the country. One need not look further than recent DHS Data Mining reports, which address the development of the Automated Targeting System-Persons (ATS-P) module.[76] The report explains, in general language, how this system is put to use; various governmental databases (also those which include personal information) are analyzed to generate predictions regarding the risks associated with those striving to cross borders.[77]

[72] *See Secure Flight Program*, TRANS. SEC. ADMIN. (May 14, 2013), http://www.tsa.gov/stakeholders/secure-flight-program.

[73] *Secure Flight*, EPIC.ORG, http://epic.org/privacy/airtravel/secureflight.html (last visited May 20, 2013) (outlining delays and how the project is moving forward, and listing additional sources).

[74] The nature of this list and the ability to seek redress and learn why someone is included in it, and how one can get off the list, received recent academic interest. *See, e.g.*, Ramasastry, *supra* note 11, at 779–92; Solove, *supra* note 12, at 344. A recent report indicates that the list is shorter than people think, including about 10,000 names, less than ten percent of whom are American. Jamie Tarabay, *The No Fly List: FBI Says It's Smaller Than You Think*, NPR (Jan. 26, 2011, 12:01 AM), http://www.npr.org/2011/01/26/133187841/the-no-fly-list-fbi-says-its-smaller-than-you-think.

[75] For a detailed discussion, see Ramasastry, *supra* note 11, at 779–92.

[76] *See Privacy & FOIA Reports*, U.S. DEPT. OF HOMELAND SECURITY, http://www.dhs.gov/privacy-foia-reports (last visited May 20, 2013) (listing all the recent data mining reports).

[77] DHS has indicated that data mining is not yet used, but might subsequently be applied for this objective. U.S. DEP'T OF HOMELAND SECURITY PRIVACY OFFICE, 2010 DATA MINING REPORT TO CONGRESS 13 (2010) [hereinafter 2010 DATA MINING REPORT], *available at* http://www.dhs.gov/xlibrary/assets/privacy/2010-dhs-data-mining-report.pdf. Indeed, the 2011 Data Mining Report already notes that "data mining queries of data in ATS and its source databases may subsequently be used by analysts to refine or further focus those rules to improve the effectiveness of their application." U.S. DEP'T OF HOMELAND SECURITY PRIVACY OFFICE, 2011 DATA MINING REPORT TO CONGRESS 11–12 (2012) [hereinafter 2011 DATA MINING REPORT], *available at* http://www.dhs.gov/xlibrary/assets/privacy/dhsprivacy_2011dataminingreport.pdf. A possible reason as to why predictive data mining is featured in the ATS-P context but not with regard to "Secure Flight" and the "No Fly List" is that only the former process is followed by human review, which allows for correcting automated errors. E-mails exchanged between Tal Zarsky and Paul Rosenzweig, Founder, Red Branch Law and Consulting, PLLC (Sept. 24, 2012, 9:57 AM) (on file with author). Note, however, that in the 2011 DHS Data Mining Report, DHS states that "TSA's Secure Flight Program (Secure Flight) began leveraging ATS-P to identify individuals requiring enhanced screening prior to boarding an aircraft." 2011 DHS DATA MINING REPORT, *supra* note 77, at 5. As ATS-P makes use of predictive data mining, this statement might indicate that data-mining measures are migrating to the "Secure Flight" context as well.

Beyond the ongoing battle against terrorism,[78] similar predictive practices are moving into law enforcement. Predictive modeling is applied to decisions regarding the allocation of resources (such as police officers, cameras, or cars) or decisions as to which individuals should be stopped for questioning.[79] It has long been used for examining forms of "white collar" crimes, such as insider trading or money laundering.[80] New programs are now put in place to further enhance these abilities.[81] In addition, DHS is moving its efforts outside of airports to track vehicles crossing borders, while applying some forms of predictive data mining as well.[82]

The outcome of being "flagged" by these mechanisms varies. It probably rarely leads to actual arrests and limitations of freedom,

[78] Several experts have questioned the ability of data mining to provide a helpful tool in this context. *See, e.g.*, Jeff Jonas & Jim Harper, *Effective Counterterrorism and the Limited Role of Predictive Data Mining*, (POL'Y ANALYSIS, No. 584, Dec. 11, 2006) *available at* http://www.cato.org/sites/cato.org/files/pubs/pdf/pa584.pdf; Bruce Schneier, *Why Data Mining Won't Stop Terror*, WIRED (Mar. 9, 2006), http://www.wired.com/politics/security/commentary/securitymatters/2006/03/70357?currentPage=all; *see also* COMMITTEE ON TECHNICAL AND PRIVACY DIMENSIONS OF INFORMATION FOR TERRORISM PREVENTION AND OTHER NATIONAL GOALS ET AL., PROTECTING INDIVIDUAL PRIVACY IN THE STRUGGLE AGAINST TERRORISTS: A FRAMEWORK FOR PROGRAM ASSESSMENT 185 (2008), *available at* http://www.nap.edu/catalog.php?record_id=12452#toc (finding these practices neither feasible nor desirable). This Article, however, assumes prediction models will be applied and examines when and how transparency is mandated to mitigate other concerns, while believing that in some contexts, prediction models can prove effective and efficient.

[79] *See, e.g.*, Press Release, IBM, Memphis Police Department Reduces Crime Rates with IBM Predictive Analytics Software (July 21, 2010), *available at* http://www-03.ibm.com/press/us/en/pressrelease/32169.wss. In a broader context, the government is considering the use of "fusion centers" to battle organized crime and drug trafficking. According to the DOJ Privacy Impact Assessment, such operations will include the use of data mining at one point. *See* U.S. DEPT. OF JUSTICE, PRIVACY IMPACT ASSESSMENT FOR THE ORGANIZED CRIME DRUG ENFORCEMENT TASK FORCE FUSION CENTER AND INTERNATIONAL ORGANIZED CRIME INTELLIGENCE AND OPERATIONS CENTER SYSTEM 9 (2009), *available at* http://www.justice.gov/opcl/crime-taskforce.pdf. For a discussion of the problems and shortcomings of fusion centers in general, see Danielle Keats Citron & Frank Pasquale, *Network Accountability for the Domestic Intelligence Apparatus*, 62 HASTINGS L.J. 1441 (2011). There are still other examples for the uses of predictive data mining in the criminal justice context. For a discussion as to how data mining is being considered for "[p]roviding well-tailored rehabilitation services to juvenile offenders" and the problematic implications of using data mining to predict recidivism, see Danielle Citron, *Data Mining for Juvenile Offenders*, CONCURRING OPINIONS (Apr. 21, 2010, 3:56PM), http://www.concurringopinions.com/archives/2010/04/data-mining-for-juvenile-offenders.html.

[80] K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 1, 15–17 n.43 (2003); GAO REPORT, *supra* note 50, at 4.

[81] An example of a new program is the Immigration and Customs Enforcement's (ICE), use of DARTTS—another system of data mining used by DHS which is addressed in the DHS 2010 Data Mining Report to Congress. 2010 DATA MINING REPORT, *supra* note 77, at 16. For a discussion of the expanding use of this system, see 2011 DATA MINING REPORT, *supra* note 77, at 17–24.

[82] The 2011 DHS Data Mining Report addresses the new ATS–Land Module (ATS–L), by stating: "ATS-L processes vehicle and occupant information against other information available to ATS, and applies rules developed by subject matter experts (officers and agents drawing upon years of experience reviewing historical trends and current threat assessments), system learning rules (rules resulting from the system weighting positive and negative results from subject matter expert rules), or affiliate rules (derived from data establishing an association with a known violator)." 2011 DATA MINING REPORT, *supra* note 77, at 13.

although it could encumber the movement of those selected.[83] Often, selection results in focusing the attention of the relevant agent and additional scrutiny being applied, at times even without the individuals' knowledge. A variety of harms, however, might be at play—such as harms to autonomy, lack of due process, and even "mere" annoyance and inconvenience. The inherent opacity of these processes further exacerbates these harms and concerns. Therefore, it is crucial to understand how transparency could be enabled in these processes.

C. Automated Prediction: Technology, the Human Touch, and Policy Decisions

The process of predictive modeling, which is premised upon personal information, includes several important elements. To enable our analysis, some initial sorting and introduction is required, and we will therefore separate these elements into three distinctive themes which we now address separately: *technology*, *human discretion*, and *policy*.

The key *technological* elements enabling this process are data-mining tools and protocols.[84] Data mining refers to both "subject-based"[85] and "pattern-based" searches.[86] This Article focuses on the pattern-based searches (also referred to as "event-based" data mining). Popular methods of such data mining are clustering, link analysis, and the construction of decision trees.[87] As its name indicates, pattern-based analysis strives to find rules and associations in data sets. To demonstrate in the context of the two examples detailed above, rather than being driven by a specific individual who generates interest or suspicion, the analysis focuses on analyzing crucial events (such as

[83] "The results of queries in ATS-P are designed to signal to [U.S. Customs and Border Protection] officers that further inspection of a person may be warranted . . . ." *Id.* at 12. Yet there might be more severe cases. According to EPIC, CAPPS II included a category of very high-risk passengers, who are to be delayed and possibly arrested. *Passenger Profiling*, EPIC.ORG, http://epic.org/privacy/airtravel/profiling.html (last visited May 20, 2013).

[84] For this discussion, I revert to a somewhat technical definition of data mining: the "non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"—yet as I will explain (when addressing the notion of "interpretable" data mining processes), the final segment of this definition is probably open to debate. Usama M. Fayyad et al., *From Data Mining To Knowledge Discovery: An Overview*, *in* ADVANCES IN KNOWLEDGE DISCOVERY AND DATA MINING 1, 6 (Usama M. Fayyad et al. eds., 1996). A somewhat different definition from a congressional report is "the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets." CRS REPORT, *supra* note 62, at 1. The GAO used yet a somewhat different definition: "we define data mining as the application of database technology and techniques—such as statistical analysis and modeling—to uncover hidden patterns and subtle relationships in data and to infer rules that allow for the prediction of future results." GAO REPORT, *supra* note 50, at 1.

[85] These pertain to database searches of and for specific individuals, events and predetermined patterns. For instance, these are the searches of matching specific names in the "Safe Flight" list to those on the governmental watch list.

[86] For a discussion regarding the distinction between the two, see Cate, *supra* note 13, at 438–39; Slobogin, *supra* note 11, at 322–23.

[87] Zarsky, *supra* note 12, at 9–15.

previous attacks or tax avoidances). The analysis strives to identify patterns that describe events and the links among them. Thereafter and while using these events, the analysis could later provide clues to locate individuals connected to other (either past or future) similar events (again, acts of tax evasion, terrorist attacks, or other subjects of law enforcement's interest). While engaging in such analysis, analysts are required to define specific parameters, and thereafter the software itself sifts through the data and points to trends within it.

Data mining is the last segment of a broader complicated and tedious task. For data mining to commence, a database must be formulated, at times by bringing together information from different sources. This is the process of data matching,[88] data warehousing,[89] and data cleansing.[90] Every one of these steps generates specific questions and difficulties—some of which pertain to transparency and will be discussed below.

While the automated nature of this technological process generates fascination, *human discretion* plays an important role in it as well. Analysts carry out extensive tasks[91] and have ample opportunity to leave an ideological (and potentially hidden) impression on the process.[92] These opportunities are evident from the very start of the prediction process; the dataset must be actively constructed, at times by bringing together data from various sources. This task requires various decisions, such as which databases should be used. Other decisions are more subtle, such as what counts as an "event" which will trigger further analysis.[93] Next, the analysts play an active role in defining the parameters of the actual data-mining analysis and the creation of clusters, links, and decision trees which are later applied.[94] This is done both in advance, and after the fact, by weeding out results the analyst might consider as random, wrong, or insignificant.

[88] In this context, Ramasastry addresses the concern that the information incorporated from various sources into one central database was collected in very different contexts and thus, when aggregated, will lose its context and validity. Ramasastry, *supra* note 11, at 761–62. This, too, is a challenge the analysts must deal with within the aggregation process.

[89] PETER CABENA ET AL., DISCOVERING DATA MINING: FROM CONCEPT TO IMPLEMENTATION 18–21 (1998); Zarsky, *supra* note 12, at 7–8.

[90] For a discussion of the process of cleaning databases from errors see Zarsky, *supra* note 12, at 7–8.

[91] Bamberger, *supra* note 21, at 672, 706–10 (explaining that, in the financial industry, automated systems are programmed to assess and limit risk).

[92] Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, *in* HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY 21, 21–39 (Batya Friedman ed., 1997).

[93] I thank Ken Farrall for illuminating this point. Ken Farrall, Address at New York University Law School: Mapping Information Flows in the Detection and Prevention of Terrorist Attacks (Oct. 13, 2010). For a discussion of the unclear definition of a terrorist event and the difficulty of defining it, see U.S. DEP'T OF JUSTICE, FEDERAL BUREAU OF INVESTIGATION, TERRORISM 2002-2005 iv-v, *available at* http://www.fbi.gov/stats-services/publications/terrorism-2002-2005/terror02_05.pdf.

[94] For more on these tools, see Zarsky, *supra* note 12, at 10 n.30, 15–16, 28.

The final stage of the data-mining process—generating a predictive model to be applied to future events and individuals—includes many opportunities for exercising human discretion as well.[95] Here, analysts must decide how to match rules premised upon existing data to the novel factors transpiring in the field. For instance, should data related to currency be adjusted for inflation? Is paying with PayPal the same as a cash-based transaction? Hundreds of such substantial decisions must be made on a daily basis.

Finally, applying these models calls for several subtle yet important *policy decisions* which are rarely made public. For instance, note the setting of the acceptable level of *false negatives* in the predictive process. As explained above, false negatives refer to the inability of the data-mining analysis to correctly reveal instances in which the sought event (such as the tax evasion) transpires. They result from a broad and diverse mix of factors and are very difficult to establish.[96] They surely must be tested, however, to assess the sustainability of the prediction model. Crossing the accepted level of such errors might ultimately lead to terminating the project—a hidden policy decision this process entails.

Another more subtle policy decision concerns *interpretation*. Thus far, we have described data mining as a process which reveals mere correlations. Data mining might point to individuals and events, indicating elevated risk, without telling us *why* they were selected. The extent to which the predictive process is understandable to humans is not accidental, but results from an important policy decision, which is often hidden and rarely discussed in policy debates. In technical jargon, this decision relates to the question as to whether the prediction process is *interpretable* or *non-interpretable*. A non-interpretable process might follow from a data-mining analysis which is not explainable in human language. Here, the software makes its selection decisions based upon multiple variables (even thousands). In these contexts, the role of the analyst is minimized.[97] The lack of interpretation not only reflects on the role of the analysts, but also on the possible feedback to individuals affected by the data-mining process. It would be difficult for the government to provide a detailed response when asked why an individual was singled out to receive differentiated treatment by an automated

[95] *See supra* notes 91–94.

[96] Some argue this even renders the entire process more art than science. *See e.g.*, Jonas & Harper, *supra* note 78; *see also* Dean Abbott, *Data Mining: Art or Science?*, DATA MINING AND PREDICTIVE ANALYTICS BLOG (Aug. 6, 2003, 11:16 AM), http://abbottanalytics.blogspot.com/2003/08/data-mining-art-or-science.html.

[97] I was told by data mining experts that this is at times the case with face and image recognition software. Interview with Dr. Ran Wolff, Professor of Computer Science, Haifa University, in Haifa, Israel (Sept. 8, 2011) (on file with author).

recommendation system. The most the government could say is that this is what the algorithm found based on previous cases.[98]

A policy decision mandating interpretable results calls upon analysts to work through the statistical outputs received, make sure they understand their meaning, and articulate them clearly. After doing so, analysts can clearly indicate the correlations between higher risks and the personal factors (such as height, age, specific credit, or purchasing history) used in the selection process that would follow. With this information, the analyst sets up profiles based on these findings while clearly defining their parameters, and applies them to future events. While some might not associate the interpretability issue with transparency, these two issues are closely linked. At times, transparency calls for proactively producing information so as to inform the public[99]—a call which could result in mandating an interpretable process.

Interpretability allows the analyst to go beyond correlation and search for a theory that could uncover *causation*. For instance, one way, cash-only airline tickets can be theoretically linked to terrorists planning to ignite explosives on an aircraft. Other correlations might call for more elaborate theories, experiments, or analyses to figure out the causation in play. When seeking correlations, analysts could be called to ignore findings which seem ridiculous or cannot be explained by a reasonable causation model. Thus, interpretability could be considered as an important step to assure the prediction process's quality, precision, and that the results it provides are not merely anecdotal.[100] Yet interpretability has a flip side as well. Mandating interpretability might render the process less complex and therefore less accurate.[101]

In addition, interpretability and causation allow analysts to identify instances where the patterns used amount to illegal discrimination. This will be the case when the patterns formulated are proxies for reliance on personal attributes which are unethical or even illegal, such as race, ethnicity, religion, and other selected factors. The concern that

[98] This is mostly the case when more advanced tools of data mining are applied, such as decision tree learning. Since these tools generate specific concerns of their own, they will not be further addressed here. For a discussion of such instances that at times involved tens of thousands of factors, see David Martens & Foster Provost, *Explaining Documents' Classification* 2 (N.Y.U.-Stern Sch. of Bus., Working Paper No. CeDER-11-01), *available at* http://archive.nyu.edu/handle/2451/29918.

[99] Schauer, *supra* note 18, at 1344.

[100] "Building" a theoretical justification for a statistical correlation, however, is usually easy and merely requires some imagination. Thus, assuring causation exists will probably prove to be a weak form of protection against inaccurate analyses. For a discussion in a very different context, see S. Stanley Young et al., Comment, *Cereal-Induced Gender Selection? Most Likely a Multiple Testing False Positive*, 276 PROC. ROYAL BIOL. SOC'Y 1211, 1212 (2009), *available at* http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2660953/pdf/rspb20081405.pdf (using the term "Retrospective Rationalization").

[101] Martens & Provost, *supra* note 98, at 5–10.

automated prediction can lead to illegal discrimination might call for an additional layer of policy steps which relate to transparency.

III. THE NATURE OF TRANSPARENCY IN PREDICTIVE MODELING: WORKING THROUGH THE INFORMATION FLOW

A. Transparency—Segment by Segment

A call for transparency can refer to a variety of segments throughout the prediction process. Assuring transparency at every segment generates specific forms of costs and balances and is derived from different laws and justifications. In some instances, transparency might merely require uploading information and disseminating it. In others, it may call for the creation of guidelines and protocols. In the most extreme cases, transparency might call for proactive research on behalf of the government.[102]

Therefore, calling for "transparency" in the context of automated prediction is overbroad and ultimately ineffective. The process must be broken up into several segments, each of which will be discussed separately. To effectively illustrate this point, this section identifies three distinct segments of the prediction process. Understanding the different challenges of every segment is the key to resolving the apparent tension between transparency and prediction schemes. These include: (1) the collection of data and aggregation of datasets, (2) data analysis, and (3) actual strategies and practices for using the predictive models, effectiveness of which could be measured by both the way they are applied ex ante and their final impact ex post.[103] This final element is established through a feedback process which follows the use of the predictive models. The discussion proceeds along two parallel lines—it addresses the forms of "transparency" required by law today and compares them with other layers of transparency which would ideally be set in place. In the context of predictive modeling, transparency is indeed already mandated by several layers of laws and regulations. Yet these requirements usually fall short of meeting the actual transparency needs called for. In addition, various laws might be pulling in different policy-related directions. This analysis strives to provide an overall

[102] Schauer, *supra* note 18, at 1344–45.

[103] I am unaware of other scholarship mapping out transparency in accordance with the analysis process in similar fashion. Recently, Julie Cohen offers a somewhat different mapping for forms of transparency. "Operational transparency encompasses transparency about the design and implementation of surveillance practices, transparency about the operation of the network's borders and flows, and transparency about the processes by which network standards are designed and adopted." COHEN, *supra* note 12, at 235. I assume that the first segment in Cohen's analysis addresses the collection (or surveillance) stage. The second element includes various segments in the process which were not defined here, and the third element includes part of both segments (b) ("analysis") and (c) ("use").

coherent perspective of where transparency stands and where it ought to be in this unique context.

Addressing the various transparency requirements already mandated by law in the governmental automated prediction context calls for an introduction to a complex regulatory "alphabet soup." At first, FOIA facilitates the disclosure of governmental operations.[104] The Privacy Act generates transparency requirements as well for many predictive practices when these are premised (as the practices discussed in this Article are) on personal information.[105] Complying with the Privacy Act allows individuals to review information pertaining to them within the relevant datasets.[106] In addition, it calls for publishing System of Records Notices (SORNs).[107] These notices provide information concerning the forms of data collected and how it is stored, disseminated, and deleted.[108] The Data Mining Reporting Act is another piece of relevant legislation. This specific law shines the light of transparency, by requiring the publication of relevant reports when data mining is put to use.[109]

Beyond these laws, important tools for providing transparency are Privacy Impact Assessments (PIAs).[110] The E-Government Act and the DHS Act mandate the publication of PIAs.[111] The call for PIAs comes with a distinct set of guidelines and practices. To demonstrate the complexities and shortcomings of this regulatory measure (as well as examining the PIA template), this study focuses on a PIA authored by the DHS to address the ATS-P project mentioned above (a project which applies data mining and possibly predictive modeling). Using the PIA as a measure to promote privacy and transparency has several known shortcomings. For instance, the process of drafting and publishing is not always conducted openly and is met by low levels of compliance.[112] Yet even when assuming these problems are temporary, this Article demonstrates central and inherent shortcomings of using PIAs. PIAs are not focusing on the full scope of information which transparency in this context calls for.

[104] *See* Freedom of Information Act, 5 U.S.C. § 552 (2006).

[105] *See* The Privacy Act of 19745 U.S.C. § 552a 2006.

[106] *Id.* at § 552a(d).

[107] *Id.* at § 552a(a)(5), (e).

[108] For instance, see U.S. DEPT. OF HOMELAND SECURITY PRIVACY OFFICE, ANNUAL REPORT TO CONGRESS: JULY 2007–JULY 2008, at 9 (2008), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_annual_2008.pdf.

[109] This Act calls upon relevant entities (such as DHS) to provide Congress with reports regarding data-mining activities. *See* Federal Agency Data Mining Reporting Act of 2007 42 U.S.C. § 2000ee-3(c) (Supp. III 2007). Additional reports are filed by the GAO and CRS, yet these are sporadic in nature and therefore only mentioned in passing below.

[110] *Privacy Compliance*, U.S. DEPT. OF HOMELAND SECURITY, http://www.dhs.gov/privacy-compliance (last visited May 20, 2013) (defining what are PIAs).

[111] *See* 44 U.S.C. § 3501 (2006); 6 U.S.C. § 142 (2006).

[112] Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy Decision Making in Administrative Agencies*, 75 CHI. L. REV. 75, 81–82 (2008).

Finally, transparency might also be facilitated by additional reports prepared in accordance with other legal requirements. For instance, the DHS Civil Rights and Civil Liberties office is called at times to file Civil Rights Impact Assessments (CLIAs).[113] The analysis below will touch upon the important role this report can play in providing transparency to the opaque and complex process of the automated prediction of human actions.

B. The Three Segments of Information Flow and Transparency

Transparency concerns arise at the first steps of the predictive modeling process—(a) *the collection of data and aggregation of datasets*. Here, transparency refers to providing information regarding the kinds and forms of data and databases used in the analysis. On its face, such disclosures generate limited social risks, which might arise when specific secretive governmental datasets are applied. In these cases, this disclosure requirement could be eased.[114]

This most basic requirement is often recognized by the existing legal framework and even met in practice. For example, the Office of Management and Budget's (OMB) Guidelines[115] for PIAs submitted in accordance with the E-Government Act[116] call for these forms of disclosure. Such information is also required in accordance with the Federal Agency Data Mining Reporting Act.[117] Indeed, the DHS discloses this information in its PIAs,[118] while drawing out the general database sources which it uses.

An additional layer of transparency at this juncture pertains to the human decisions made during the aggregation and collation stage. Human discretion plays out in a broad array of crucial stages in this specific context. For instance, humans must decide as to the way similar

[113] *See Civil Rights & Civil Liberties Impact Assessments*, U.S. DEPT. OF HOMELAND SECURITY, http://www.dhs.gov/civil-rights-civil-liberties-impact-assessments (last visited May 20, 2013) [hereinafter CLIA Website].

[114] Even in the IRS example this aspect is of relevance, as disclosure might pertain to secret sources, such as the use of informants, which the IRS will not disclose. *See supra* note 58.

[115] OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, OMB MEMORANDA NO. M-03-22, OMB GUIDANCE FOR IMPLEMENTING THE PRIVACY PROVISIONS OF THE E-GOVERNMENT ACT OF 2002, Attachment A § II(C)(a)(1)(1)(2003) [hereinafter OMB GUIDANCE] *available at* http://www.whitehouse.gov/omb/memoranda_m03-22#a ("PIAs must analyze and describe. . . what information is to be collected (e.g., nature and source). . . .").

[116] 44 U.S.C. § 3501.

[117] 42 U.S.C. § 2000ee-3(c)(2) (2006).

[118] For instance, see U.S. DEP'T OF HOMELAND SECURITY, PRIVACY IMPACT ASSESSMENT FOR THE AUTOMATED TARGETING SYSTEM 6–8 (2007) [hereinafter ATS-P PIA], *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_cbp_ats_updated_fr.pdf. As mentioned, PIAs prepared by the DHS are subject to an additional set of guidelines. These guidelines require forms of disclosure similar to those addressed in the Data Mining Reporting Act. *Id.* at 6–13. These guidelines even go into additional detail so as to include specific factors which must be addressed in the PIA report. *Id.* at 6–7.

records in different datasets are matched into one source.[119] These matters are not addressed in the PIAs. Such transparency measures do receive limited attention and treatment in the Computer Matching and Privacy Protection Act, however, which provides for some accountability and limited transparency for these actions.[120]

To fully and properly meet transparency requirements at this juncture, access should be provided to the working protocols analysts use for these early segments of the prediction tasks. This is easier said than done. Clear protocols regarding the human role in data aggregation might not exist. Therefore, transparency will call for their creation, updating, and enforcement—elements currently not covered by the law.

Finally, transparency in this early stage of the analysis has an additional, more extensive meaning. It might call for presenting the actual data used in the analysis process. In some contexts, a right to review such data already exists. The Privacy Act provides individuals with the right to access some of the information stored about them in government records.[121] Such rights do not extend, however, to other individuals who are not data subjects.[122] Even for data subjects, such rights are rarely exercised or enforced.[123] In fact, enabling such data access rights generates additional privacy and security challenges.

Transparency considerations play a role in the next segment of the analysis process as well—(b) *data analysis*. This stage includes both technical and human-related aspects. The "technical" aspect relates to the technology used in this process. It could be rendered transparent by disclosing the names of the software applications used (if they are in commercial use). If these are custom-made, transparency could be achieved by releasing these programs' source code.

Currently, transparency of this aspect is only superficially met.[124] In PIAs, the "technology" segment usually addresses the software used—indicating if it was bought off the shelf or internally developed (an almost

[119] Studies indicate that this stage is a "major contributor to inaccuracies in data mining." Cate, *supra* note 13, at 470.

[120] This law, however, includes a law enforcement exemption. Computer Matching and Privacy Protection Act of 1988, Pub. L. No. 100-503, 102 Stat. 2507, 2512 (codified at 5 U.S.C. § 552a(a)(8)(B)(iii) (2006)); Cate, *supra* note 13, at 467. When the law applies (despite the exemption), it requires notice on data sharing and some form of overall supervision by a Data Integrity Board. *See* U.S. DEP'T. OF HOMELAND SECURITY PRIVACY OFFICE, ANNUAL REPORT TO CONGRESS: JULY 2009–JUNE 2010, at 17 (2010), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_annual_2010.pdf (outlining DHS's compliance with these requirements). In general, this requirement does not appear to enhance transparency, but merely overall accountability.

[121] 5 U.S.C. § 552a(d).

[122] *Id.*

[123] *See* Solove, *supra* note 12, at 359 (writing in a more general context). "Suppose a person disagrees with the profile. Can this be addressed at the hearing? Likely not, as the profiles are secret." *Id.*

[124] For a call to change this outcome, see Citron, *Technological Due Process*, *supra* note 30, at 1308–09.

useless piece of information).[125] The Data Mining Reporting Act addresses this issue explicitly, while calling for a *"thorough description of the data-mining technology that is being used or will be used*."[126] Compliance with this segment is limited, however. The DHS 2010 data mining report, for instance, indicates that several data-mining ventures use "commercial off-the-shelf" software (COTS)[127] but does not mention which software is used. In some instances, the report notes the use of "custom-designed software," with no additional information.[128] As with the IRS example, it appears that a lapse in the theory regarding the meaning and importance of transparency at this juncture allows the practical/regulatory practices to sidestep the existing transparency requirements.

On the human side, transparency requirements might pertain to a variety of elements. In many automated predictive processes, analysts are required to establish a sufficient level of *support* and *confidence*.[129] First, analysts must establish the level of "support" which would be acceptable for the "rules" uncovered; namely, how frequent or obscure the uncovered pattern can be within the database so as to be further considered. A correlation of events could be very interesting indeed, yet if it occurred very few times, (and thus with limited "support") it is probably statistically insignificant (for instance, an analysis of tax returns might find a trend of data indicating that law professors who write about both transparency and e-commerce overstate their charity exemption—a pattern which pertains to too few taxpayers). In addition, limited support might indicate a finding that is statistically significant, yet still rare and thus not worth applying in practice (given the costs of implementation). The analysts must also establish an accepted "confidence" level. This factor refers to the degree of accuracy of the rule produced. It relates to the level of "false positives" in the process (for instance, if again the analysis reveals a pattern of the abovementioned law professors overstating the specific exemption, but only in twenty-five percent of the cases—is that a high enough chance to act upon?).

[125] *See, e.g.*, ATS-P PIA, *supra* note 118. I assume such language and form of disclosures are set forth in view of specific requirements in the OMB guidelines, explaining that a PIA must "identify what choices the agency made regarding an IT system or collection of information as a result of performing the PIA." OMB GUIDANCE, *supra* note 115, at Attachment A § (II)(C)(a)(2). I, of course, do not find the disclosure form addressed in the text either informing or suitable.

[126] 42 U.S.C. § 2000ee-3(c)(2)(B) (2006) (emphasis added).

[127] 2010 DATA MINING REPORT, *supra* note 77, at 10.

[128] *Id.*

[129] According to a data mining expert, the use of these terms lacks precision. When used in this context, the term "confidence" is merely a "quick and dirty" shortcut to describe the difficulties of data mining, and the term "support" does not carry the same meaning used in the statistical context. Wolff, *supra* note 97. More sophisticated statistical elements are required to present these difficulties in detail. In the context of this Article, however, these basic elements should do.

In existing laws, some of these decisions are already addressed. The Federal Data Mining Reporting Act calls for disclosures,[130] which seem to coincide with the "support" and "confidence" elements mentioned above, by calling for disclosure regarding the "*basis*" for relying upon the data-mining findings.[131] Yet again, it appears that the legal framework is too broad and is failing to produce meaningful results in practice. The disclosures made thus far to meet this Act's requirements do not address the factors noted above. They merely generally address the tasks analysts face and the prospects for their success.[132] A better theoretical understanding of what stands behind such requirements could hopefully improve this outcome.[133]

Intuitively, when addressing transparency in this specific context, public opinion and academic inquiries usually focus on the third ("C"), *usage stage* in which the predictive patterns are utilized. This stage generates several transparency concerns. First, transparency discussions at this juncture call for the disclosure of the actual strategies and practices for using the data. In other words, these are the *predictive models* formulated through the data-mining process. They are the actual "profiles," according to which the IRS or other entities single out individuals or events.

Governments are reluctant to provide transparency at this juncture, and such reluctance is mirrored in the existing legal rules. The IRS does not provide such information, even to the relevant taxpayer. Case law indicates law enforcement is not required to make such disclosures

[130] 42 U.S.C. § 2000ee-3(c)(2)(B). The full language is: "including the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity." *Id.*

[131] The OMB Guidelines do not address this issue directly. *See* OMB GUIDANCE, *supra* note 115.

[132] ATS-P PIA, *supra* note 118, at 4, 11, 28.

[133] The analysis in this Section assumes that all the factors here discussed pertain to disclosure regarding final decisions made by government (or can be referred to as "post-decisional" step). In doing so, our discussion sidesteps the thorny issue as to possible requirements for disclosing internal governmental deliberations, protocols, discussions, and memos. These are usually considered outside our discussion of transparency requirements, and covered (thus rendered opaque) by the "deliberative process privilege," which is also somewhat accepted into the FOIA in exemption 5. For more on this privilege, see *Freedom of Information Act Guide: Exemption* 5, U.S. DEP'T. OF JUSTICE (May 2004), http://www.justice.gov/oip/exemption5.htm. One can here argue that the abovementioned assertion is misplaced, and that the Analysis portion of the automated predictive process in its entirety is pre-decisional by nature—these are steps that come before the final "decision"—which is the nature of the profile or pattern being used. *See infra* Table 1. If this final assertion is correct, the Analysis stage should be rendered opaque in its entirety as well, and the transparency measures considered in the text should be set aside. I find this pro-opacity argument unconvincing. While the structuring of the patterns could be seen as a process, it is one that would be ongoing and premised upon standards and technologies which were previously set by the government. In other words, our discussion here addresses disclosures of "meta" rules, which establish the nature of the automated prediction, rather than the actual ongoing internal discussions between employees implementing the prediction process, which should remain privileged.

according to FOIA.[134] Furthermore, the concept of "law enforcement" in this context is defined broadly to include any attempts to uphold regulation.[135]

In the context of border protection and anti-terrorism, opacity in this context prevails as well. The PIAs addressing these actions are not providing a sufficient response to these transparency concerns. Again, we can witness here a broad disparity between the language of the regulation and its actual implementation. GAO Guidelines for structuring PIAs call for presenting the consequences of information collection and flow.[136] Thus, on its face, a response to this requirement would fulfill the transparency concerns arising at this juncture. Yet this obligation is instead fulfilled while using very general language (in the same way the IRS responded to transparency requirements). The DHS Privacy Guidelines address this issue as well, but with a much broader brush. They call for informing the public of the use of such data mining.[137] Again, this requirement is fulfilled with broad and empty disclosures. In the ATS-P PIA, information regarding these factors is not even provided. Furthermore, this PIA specifically notes that the government need not provide an individual with information regarding her or his "assessment" under the Privacy Act.[138] To conclude, the current understanding and implementation of existing regulation are leaving individuals in the dark regarding this crucial piece of the automated prediction process.

Beyond the current regulatory failure, there is a practical problem lurking in the background which has yet to be discussed. Given the technical nature of this process, regulation cannot merely call for disclosing the factors used in a profiling scheme. Data mining is a dynamic, ever-changing, and at times, automated process. Therefore, with advanced prediction, there might be *no static "profile" to reveal*. There is merely a dynamic learning process enabled by computer algorithms. The algorithm might rely upon a complex association rule which includes a multitude of factors, as well as the interaction among them. It might also rely upon clusters with blurry and constantly

[134] *See* discussion *supra* note 54. It is generally recognized that the Privacy Act is not an effective tool for these disclosure objectives. *See* Ramasastry, *supra* note 11, at 793. For a discussion of FOIA's exemptions, see COHEN, *supra* note 12, at 209.

[135] *See, e.g.*, Gerstein v. CIA, No. C-06-4643, 2010 WL 669743, at *5 (N.D. Cal. Feb. 23, 2010).

[136] OMB GUIDANCE, *supra* note 115, at Attachment A § II (C)(a)(1).

[137] DHS GUIDELINES, *supra* note 23, at 19. These guidelines are implemented, and the ATS-P PIA report indicates the usage of data mining schemes. ATS-P PIA, *supra* note 118, at 11–12. These elements are important, yet our analysis strives to go far beyond these disclosure requirements and seeks transparency in the intricacies of the data mining process, which this statement does not seem to answer to.

[138] ATS-P PIA, *supra* note 118, at 23 ("With respect to the data that ATS creates, i.e., the risk assessment for an individual, the risk assessment is for official law enforcement use only and is not communicated outside of [U.S. Customs and Border Protection] staff, nor is it subject to access under the Privacy Act.").

changing borders. Conveying information about these practices to the public in an understandable form calls for setting new regulatory paradigms in place.[139] A more radical view may even require all relevant processes to be interpretable and thus enable simple and understandable disclosures to the broad public.

Not only might transparency impact the way the prediction process would unfold, it might call upon government to produce new sets of information.[140] The government might not only be required to present the factors it strives to use in the prediction process, but also the causation theory behind their selection. Furthermore, the government might be required to assure that the prediction scheme does not involve the use of factors that are considered off limits (either directly, or by proxy) because they are discriminatory and unethical. Today's legal setting is a far cry from achieving these objectives.

Case law (and a recent Department of Justice (DOJ) memorandum) indicates that neither FOIA nor the Privacy Act calls upon the government to generate *new* information, but to merely disclose existing information.[141] Here too, new regulatory measures (and theoretical steps that back them) are called for. Finally, unique transparency requirements relate to examining the usage of prediction models from an additional important perspective—after the fact. From this perspective, examining the use of predictive models can lead to important insights. It reveals how many of those whom the governmental prediction models indicated as a higher risk (for tax fraud, terrorist activities, or other felonies) indeed turned out to be of no risk at all (and thus are considered as "false positives"). It could further indicate how many of those considered as lower risks should have been indicated as a high risk, yet were missed (and were thus "false negatives"). In addition, the analysis of the ongoing process from an *ex post* perspective will provide information whether the practices

[139] Note that the regulation of credit analysis in the commercial sector has made some advances in this realm. Here, according to the FCRA, an individual should be provided with the reasons for his or her credit result—more specifically the four key factors which lead to this result. For a discussion of this system and its flaws, see *Frank Pasquale*, THE BLACK BOX SOCIETY: PRIVATE TECHNOLOGY AND PUBLIC LIFE (forthcoming 2013) (manuscript on file with author).

[140] FUNG ET AL., *supra* note 24, at 41–43 (noting that at times disclosure calls for producing additional information).

[141] *Overview of the Privacy Act of 1974*, U.S. DEP'T. OF JUSTICE (2010) ("It should be noted that the Privacy Act—like the FOIA—does not require agencies to create records that do not exist. *See* DeBold v. Stimson, 735 F.2d 1037, 1041 (7th Cir. 1984); Harter v. IRS, No. 02-00325, 2002 WL 31689533, at *5 (D. Haw. Oct. 16, 2002). . . . *But compare* May v. Dep't of the Air Force, 777 F.2d 1012, 1015–17 (5th Cir. 1985) ('reasonable segregation requirement' obligates agency to create and release typewritten version of handwritten evaluation forms so as not to reveal identity of evaluator under exemption (k)(7)), *with* Church of Scientology W. United States v. IRS, No. CV-89-5894, slip op. at 4 (C.D. Cal. Mar. 5, 1991) (FOIA decision rejecting argument based upon *May* and holding that agency not required to create records).").

facilitated *de facto* illegal or unethical forms of discrimination.[142] This step of the process is crucial in an automated predictive process, where steps are premised on statistical analyses and various assumptions which might be totally off mark yet are difficult to challenge.[143]

The *ex post* perspective of prediction processes is therefore a crucial piece of the overall transparency puzzle, yet is often overlooked by academics and policy makers.[144] The foundations for meeting these needs are already in place, however. The Office of Civil Rights and Civil Liberties within the DHS is at times called to issue Civil Liberties Impact Assessments (CLIAs).[145] According to the existing template for such reports,[146] they must examine how new programs and policies will (among others) affect minorities. They also must examine what alternative routes could be taken to meet the same objectives while limiting harm to civil liberties.[147] In other contexts, however, new rules would be required to generate and later publish these "*ex post*" studies.[148] Such studies will allow experts to establish whether the human traits examined are predictable along time and across people.

To be fair, there are additional transparency-related concerns that the mentioned regulatory frameworks indeed address and this Article

[142] For instance, see the results of a study concerning NYPD policy for stopping individuals, which turned out to be extremely biased. *See* Second Supplemental Report of Jeffery Fagan, Ph.D., Floyd v. City of New York, 08 Civ. 01034 (SAS) (S.D.N.Y. Dec. 12, 2012), *available at* http://ccrjustice.org/files/FaganSecondSupplementalReport.pdf.

[143] One such assumption is that the information is "stationary"—namely, that the dataset used for the analysis is actually relevant to the present and future—and therefore trends identified in the past are still of relevance.

[144] In a recent draft, Shkabatur explains that "performance transparency" (which she considers as information regarding how government policy was implemented and how successful the policy is in achieving its objectives) should be considered as part of an overall transparency framework. Jennifer Shkabatur, *Transparency With(out) Accountability: Open Government in the United States*, 31 YALE L. & POL'Y REV. (forthcoming 2013). Shkabatur further notes that to some extent, such disclosure is already mandated in accordance with the Government Performance and Results Act (GPRA). *Id*; *see* 5 U.S.C. § 306 (2006). The Act calls for self-evaluation on an annual basis of the agencies' ability to meet goals. Furthermore, this Act was recently amended to include additional requirements, such as posting information on a governmental website. Yet at the time of this writing, these amendments have only began to come into force. Thus it is difficult to currently judge their effectiveness. *See* GPRA Modernization Act of 2010, Pub. L. No. 111-352, 124 Stat. 3866 (2011).

[145] This is done either by law or within the agency. *See* U.S. DEP'T OF HOMELAND SECURITY, REPORT TO CONGRESS ON THE DEPARTMENT OF HOMELAND SECURITY OFFICE FOR CIVIL RIGHTS AND CIVIL LIBERTIES 20–21 (2008), *available at* http://www.dhs.gov/xlibrary/assets/crcl_annual_report_FY_2008.pdf.

[146] U.S. DEPARTMENT OF HOMELAND SECURITY OFFICE FOR CIVIL RIGHTS AND CIVIL LIBERTIES, CIVIL LIBERTIES IMPACT ASSESSMENT TEMPLATE (2008), *available at* http://www.it.ojp.gov/documents/Civil_Liberties_Impact.pdf.

[147] CLIAs have yet to examine all the aspects here addressed, but it is possible that such efforts are on their way. *See* CLIA website, *supra* note 113.

[148] Stuntz, who advocates strongly against transparency, argues that this form of information is extremely important to enable the monitoring of the actions of government; namely, more information about outcomes and less about process. William J. Stuntz, *Against Privacy and Transparency*, NEW REPUBLIC (Apr. 17, 2006), http://www.tnr.com/article/against-privacy-and-transparency. It should be noted, however, that Stuntz refers mainly to information regarding arrests and prosecutions, which are merely a small and limited segment of the form of data here discussed.

does not. Transparency requirements pertain to the steps taken to assure data security, retention, and tools for access control to the personal data used at the first two stages (collection and analysis). They further address measures for providing data accuracy and lack of errors at the collection stage and redress for harmed citizens at the usage stage. I choose to set all of these issues aside. These are important and indeed pertain to any general analysis of personal or important information. Yet predictive modeling calls for additional dimensions of disclosure which have yet to be explained and brought into an overall theoretical framework—a process I now discuss.

IV. WHY TRANSPARENCY?

A. Theories, Data Recipients, and Flow Stages

A call for transparency is echoed throughout the debate concerning the implementation of predictive data-mining tools for the analysis of personal information. The need for transparency is motivated by a variety of reasons and theories. Every one of these theories, in turn, could lead to a different solution. Many of these theories have unique implications and meanings when considered in the innovative context of automated prediction. To provide an overall taxonomy of transparency concerns and set forth solutions, we must move to map out these theories and provide an explanation as to the policy trajectory every theory implies, while accounting for the segment of the process it pertains to, and the form of transparency it calls for.

Prior to delving into a discussion of detailed theories of transparency and disclosure, we must address the simplest and perhaps most intuitive theoretical explanation. The acts of a liberal and democratic government must, categorically, be as transparent as possible.[149] Such a result is derived from the notion of democracy.[150] Scholars note that transparency is essential for democracy to function. In doing so, they make reference to an abundance of sources, drawing from Kant,[151] Locke, Mill, Rousseau, Bentham, and James Madison.[152] In a recent article, Mark Fenster summarizes the theoretical connections

[149] The Obama Administration has accepted this notion, as stated on the White House website. "Government should be transparent. Transparency promotes accountability and provides information for citizens about what their Government is doing." Transparency and Open Government, THE WHITE HOUSE, http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/ (last visited May 20, 2013) (emphasis omitted).

[150] This Article sets aside the discussion as to whether governmental transparency could be derived from a constitutional analysis. Scholars reviewing this question find it to be inconclusive. *See* Adam M. Samaha, *Government Secrecy, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 970 (2006).

[151] *Id.*

[152] Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 895–96 (2006).

between transparency and democracy,[153] noting that transparency enables an informed public debate and generates trust in and legitimacy for government. It also informs individual decision on election day.[154] A similar notion is reflected by finding transparency to be a basic human right.[155]

Accepting such a categorical argument, on its face, shortens our analysis. An instrumental (or, as recently noted by Larry Solum, a "consequentialist")[156] analysis of the benefits and outcomes of transparency as it pertains to both specific segments of society and the different steps of the process, however, is still required. A call for transparency is quickly rebutted by powerful and convincing counterarguments. Without a deeper understanding of the interests in play, properly balancing transparency against them would prove impossible. Furthermore, realistically speaking, without a clearer idea of transparency's benefits, this right will probably be conceded to other, seemingly more important, interests, as often is the case.[157] Finally, a categorical transparency right will fail to take into account many important distinctions between levels of transparency throughout the information flow addressed here. Only a broad theoretical foundation will provide specific responses at every juncture. An overarching theory of transparency rights, however, which is premised on democracy, still has an important implication; the "default" state for governmental actions should indeed be that of transparency.

The concept of transparency is too broad for an overall theoretical mapping in a single Article. Therefore, the analysis addresses four central theories of transparency which all emerge in this context: theories premised on fairness and efficiency, on policies for striving to benefit from popular innovation and crowdsourcing, on protecting information privacy rights, and on two forms of rights related to personal autonomy. In addition, the analysis addresses two central counterarguments which examine whether transparency undermines governmental initiatives and promotes (or perhaps inhibits) the development of stereotypes and stigma. Therefore, the Article brings together a diverse realm of theories

[153] *Id.* at 894. Furthermore, this notion underlies the Freedom of Information Act, which allows *any* individual to access information about the actions of government, regardless of his background or motivation. *Id.* at 898. For more on the connection between transparency and democracy, see Schauer, *supra* note 18, at 1348–50.

[154] *See* Fenster, *supra* note 152, at 898; *see also* Samaha, *supra* note 150, at 920.

[155] For a discussion of this right, from a comparative perspective, see TOBY MENDEL, FREEDOM OF INFORMATION: A COMPARATIVE LEGAL SURVEY (2d ed. 2008), *available at* http://portal.unesco.org/ci/en/files/26159/12054862803freedom_information_en.pdf/freedom_information_en.pdf.

[156] *See* Solum, *supra* note 18.

[157] *See* Bamberger & Mulligan, *supra* note 112, at 86–87 (explaining that privacy tends to lose out in policy debates to other objectives such as security, due to the fact that it has limited political capital, and that it is unlikely that politicians will risk their reputation to protect privacy). I believe it is fair to make a similar assertion with regard to transparency.

which are usually unrelated. Yet they all converge in this specific and unique context—that of automated predictions.

The previous Section demonstrated how transparency calls for varied requirements at different stages of the process. But, a final additional distinction is required prior to launching our theoretical inquiry. Transparency calls for a transfer of information. Fully understanding this concept, however, calls for distinguishing among the *recipients* of the information transparency policy provides.[158]

Intuitively, transparency is linked to merely one meaning—that the relevant information is disseminated broadly to (1) the *general public*. Some theories, however, can only justify limited exposure to (2) *institutions* which could be other branches of government, such as representatives of the judiciary[159] and the legislator (including relevant House committees)[160] or external expert panels, independent watchdog groups, or specific NGOs.

In addition, a different set of theories can justify information sharing only with (3) specific *affected individuals*. These could be the *data subjects*—individuals whose information is being analyzed by the government—or *individuals impacted* by a data-mining-driven process. In the context of predictive analysis, mapping out this final category is a challenge. While the two subgroups mentioned usually overlap, this no longer is true. One individual's data could be closely examined and used for the training set—generating a decision tree or pattern, while never actually affecting the specific individual. Other individuals, who provided limited data, would be greatly affected by this process, as it is used to distinguish among them.

To summarize, the following table lists both the transparency stages addressed above and the forms of data recipients introduced here. They will all be further discussed in the theoretical discussion that follows.

[158] For a basic discussion of this distinction, see Schauer, *supra* note 18, at 1345. Schauer addresses this factor as the "permitted audience" variable. *Id.* The text below strives to properly define the various values this factor might hold.

[159] TAPAC REPORT, *supra* note 12, at 47 (TAPAC recommendations calling for assigning the role of examining these projects to the Foreign Intelligence Surveillance Court (FISA) courts).

[160] For a recent move to expand such actions by Rep. Darrell Issa, see David M. Herszenhorn, *Chairman Seeks New Power for Watchdogs*, N.Y. TIMES, Nov. 28, 2010, at A1.

TABLE 1

Transparency Stages	**Potential Data Recipients**
A. Collection & Aggregation	1. General Public
B. Data Analysis	2. Internal and External Institutions
C. Examining Model Uses (*ex ante* and *ex post*)	3. Affected Individuals

B. Transparency—From Theory to Policy

1. Transparency as an Incentive for Fair and Efficient Policy

The most basic and popular justification for transparency is that it facilitates a check on governmental actions. These actions might be flawed, biased, ineffective, or inefficient. The relevant officials might be improperly balancing rights and interests, led by their own bigotry, or over-influenced by private interests. This outcome might result from the relevant governmental agencies' incompetence, corruption, negligence, mere error, or perhaps unacceptable point of view.[161] Officials might also try to expand their authority to meet other objectives.[162]

When discussing the specific objective that transparency might promote, the notion of accountability quickly comes to mind.[163] Transparency renders government actors accountable for their actions and their outcome. Transparency is, at times, considered synonymous to accountability. Yet these concepts clearly are not the same.[164] Accountability refers to the ethical obligation of individuals (in this case, governmental officials) to answer for their actions, possible failings, and wrongdoings. Transparency is an essential tool for facilitating

[161] Schauer, *supra* note 18, at 1349.

[162] Cate, *supra* note 13, at 485; TAPAC REPORT, *supra* note 12, at 39.

[163] *See, e.g.*, Citron & Pasquale, *supra* note 79, at 1448–55 (focusing on accountability as a measure to cure the problems generated by the growing use of Fusion Centers).

[164] For more on this distinction, see Schauer, *supra* note 18, at 1346. Scholars have noted several difficulties with meeting accountability without transparency; they note that especially when both the legislator and the executive belong to the same party, the internal balances among powers is insufficient without full transparency. DANIELE J. SOLOVE, NOTHING TO HIDE: THE FALSE TRADEOFF BETWEEN PRIVACY AND SECURITY 86 (2011) (noting violations of FISA). Furthermore, free access to all resources enables government agencies to provide better checks over each other (and in part can even benefit from public inquiries and investigations). *See* Paul M. Schwartz, *Reviving Telecommunications Surveillance Law*, 75 U. CHI. L. REV. 287, 313 n.164 (2008) (citing Neal Kumar Katyal, *Internal Separation of Powers: Checking Today's Most Dangerous Branch from Within*, 115 YALE L.J. 2314, 2341–42 (2006)). This option is also advocated by Stuntz, *supra* note 148. *But see* Antonin Scalia, *The Freedom of Information Act Has No Clothes*, REG., Mar./Apr. 1982, at 14.

accountability because it subjects politicians and bureaucrats to the public spotlight. Yet, it is merely one of the strategies that could be applied to achieve the accountability objective.[165]

A call for transparency requires the expansion of information sharing schemes beyond merely providing information to internal governmental institutions, possibly even to the broadest realm of the entire public. The assumption that broadening the scope of information sharing in this way will promote fairness and efficiency should not be taken for granted. We must examine *why* additional exposure promotes this objective, while taking into account the specific context of automated predictions.

A constructive way to approach the benefits of transparency in this context is to return to the work of Louis Brandeis. Brandeis famously advocated the use of transparency to promote fairness.[166] In a recent article, Professor Lawrence Lessig[167] drew out two basic theories as the foundation of Brandeis's call for transparency—(1) shaming, and (2) the effects of market or democratic forces.[168] These two theories are constructive in establishing the optimal breadth of transparency in the context at hand.

Transparency facilitates "shaming."[169] The fear that a broad segment of the public will learn of the bureaucrats' missteps will deter these decision makers from initially engaging in problematic conduct. Presumably, for effective shaming, the government must disseminate information to the greatest extent possible. This statement, however, relies on two hidden assumptions: (1) the public takes interest in the relevant workings of government (here, facilitating the predictive modeling dynamic), and (2) the officials and bureaucrats engaging in these practices will respond to the public's knowledge and "shaming" by

[165] Accountability could be achieved through a variety of strategies: regulation *ex ante* and by setting access controls and audit functions, which will regulate behavior *ex post*. *See* Ira S. Rubinstein et al., *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*, 75 U. CHI. L. REV. 261, 267, 269 (2008). For a concrete example of such a system, see Daniel J. Weitzner et al., *Information Accountability*, 51 COMM. ACM, No. 6, June 2008, at 85.

[166] *See* BRANDEIS *supra* note 21, at 92.

[167] Lawrence Lessig, *Against Transparency*, NEW REPUBLIC, http://www.newrepublic.com/article/books-and-arts/against-transparency (last visited May 20, 2013).

[168] Note, however, that while these tools are discussed here as measures for providing incentives for more commendable governmental actions, these arguments originally focused on the private sector. *See* BRANDEIS, *supra* note 21, at 101; *see also* Noah Feldman, *In Defense of Secrecy*, N.Y. TIMES MAG., Feb. 15, 2009, at MM11.

[169] Scholars have recently examined the return of shame-based punishment in criminal law, while pointing out various benefits and shortcomings. Dan Kahan, for instance, is a famous supporter of shaming, with limitations, although in a recent paper he expressed some reservations. *See* Dan M. Kahan, *What's* Really *Wrong with Shaming Sanctions*, 84 TEX. L. REV. 2075, 2095 (2006). I will not address this debate within the confines of this Article, mainly because this Article does not focus on using shaming as punishment (even as an alternative to imprisonment) but as a tool for generating accountability. Thus, many of the social and psychological dynamics addressed in this shaming-related literature would not apply.

adhering to the public's standards. In the case of automated predictions, both assumptions could be questioned at specific junctures of the overall process.

First, let us test the assumption that the public opinion has an interest in the specific context at hand. It should be noted that "public interest" does not necessarily call for an active ongoing interest by a broad segment of the population.[170] Information flows in cascades. The limited interest of few experts (or reporters and interest groups for that matter) can generate much greater interest by broader segments of the population. The experts encounter information, comment on it, and distribute it to the public, which picks up on these dynamics. Transparency allows for the authentication of the information the experts address. Establishing whether a shaming dynamic will transpire calls for examining whether the broad disclosure might generate an interesting story, which could be conveyed to the public. The public (directly or through proxies) might shy away from technical, complicated, and obscure matters. In such contexts, shaming might not follow.

With these insights in mind, let us examine the process's stages. Every one of them includes some issues with broader appeal, and others that are technical and complex. The collection stage (designated as A. in Table 1, *supra*) introduces seemingly salient steps such as the selection of factors for the prediction process. A reporter picking up a story as to the government collecting personal information or purchasing it on the secondary market to apply it to analyses for locating tax evaders will probably be successful in generating an interesting article. These are decisions which will create interest and uproar if deemed problematic. Decisions pertaining to the second, "analysis" stage (designated as B. in Table 1, *supra*), however, will probably generate far less interest and traction given their technical nature—consider, for example, a reporter striving to publish an article about the nature of the software used in the process. Drawing conclusions regarding the "usage" stage (designated as C. in Table 1, *supra*), which includes the process of applying models and profiles to the public, is difficult. Some of the broader issues that transparency might reveal—such as what forms of "discrimination" these models facilitate—will be sure to generate interest. Novel data-mining applications, however, rely upon technical terms. These might be too subtle to generate shame. A similar point could be made regarding the

[170] Seth F. Kreimer, *The Freedom of Information Act and the Ecology of Transparency*, 10 U. PA. J. CONST. L. 1011, 1056 (2008). Schwartz and Janger made a similar point in a somewhat different context—that of notification rules regarding data security breaches. *See* Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 MICH. L. REV. 913, 915 (2007). Schwartz and Janger explain that breach notification rules are not necessarily generating a direct market dynamic in which consumers "punish" firms, but an indirect one in which the information regarding breaches flows to the media, legislators, and regulators, and in that way have a delayed, yet substantial effect. *Id.* at 956.

ex post aspects of the usage stage. When viewing such information, the public would be interested in the overall success of the project, as well as in systematic errors and failures such *ex post* analyses reveal. Yet it might ignore the more technical aspects of the dynamics.

Furthermore, for "shaming" to have an effect, the "shamed" must respond to it. It will fail to impact, for instance, if public opinion does not associate the decision maker with the relevant action. The nature of automated prediction leads to the fact that many important decisions will be made by lower-level analysts and IT experts. Shaming might not have the needed effect on these officials. They might already be at another position by the time the information is disclosed and understood, or not clearly indicated by the disclosure. Again, shaming seems to have a limited effect on the more technical elements of this project, calling into question the wisdom of mandating transparency regarding such factors.

Finally, shaming will work well when transparency reveals official conduct that conflicts with well-established norms or existing laws. For instance, sloppily constructing and operating the data-mining process will easily generate backlash. Alternatively, it can prove valuable if transparency revealed that rather than relying on neutral factors, officials reverted to relying upon sensitive factors (either knowingly or unknowingly) such as race and religion—practices which are socially unacceptable. Yet shame might not prove helpful in other important instances, where social norms have yet to formulate. For instance, there is no social norm regarding acceptable measures of data collation and levels of support and confidence. Much of the information available in the "analysis" and "usage" stages falls within this latter category. The risks of false positives and the forms of correlations used are currently a "grey area" in terms of social norms. While transparency could be an important measure to promote a discussion on these issues,[171] it is questionable whether this context will generate shaming and therefore provide an effective check on governmental actions.

The second Brandeisian justification for transparency is related to market forces (when set in the private context); markets will punish those acting improperly.[172] Therefore, it is important for as many market participants as possible to be informed so to assume the market dynamic noted moves forward. When adapting this theory to the public context, the theory would state that disclosures of negative conduct will have political consequences. Elected officials will sanction the bureaucrats

[171] Of course one can here note that only with full disclosure regarding these notions would a social discourse concerning the acceptability of these practices arise. While I agree that public knowledge of these projects will promote such a discourse, I am not sure that full knowledge of these practices is indeed required to launch this discussion. Less aggressive means, such as informing the public in general about these practices, might suffice. Therefore, the idea that transparency is crucial for promoting a discussion regarding the development of social norms in specific areas is somewhat of a weak pro-transparency argument.

[172] *See* BRANDEIS *supra* note 21, at 101–04.

engaging in problematic practices, or force them to change course.[173] Again, this rationale calls for distributing information as broadly as possible to launch this political dynamic. The response to the TIA initiative serves as an interesting test case for this concept. Famously, politicians moved swiftly to block TIA, sensing that this is a matter close to voters' hearts.[174]

This justification of transparency relies on implicit assumptions similar to those mentioned in the shaming context—the public's interest and politicians' attentiveness to the relevant issue (which will lead to taking action against those in charge of the problematic practices). Yet, as explained above, at several crucial junctures these assumptions should not be taken for granted. At those points, transparency would be difficult to justify.

Yet this second sub-theory has some unique traits. The "market-based" perspective of this theory illuminates a concern with transparency often mentioned in the literature: the transparency of governmental actions might lead politicians to making conservative or even populist yet inefficient and possibly unfair decisions.[175] This would follow from their fear that the public will "punish" them for not meeting the public's short-term expectations, which might not even be in the public's overall interest.

Decision makers' tendency to move to meet the public's short-term expectations might lead to problematic outcomes especially when considering transparency policy for the "usage" stage (designated as C. in Table 1, *supra*) of the predictive process. The automated prediction practices here discussed are innovative by nature. The intricacies of statistics and the singling out of specific populations might prove to be political dynamite given the salience of the notion of discrimination in society and the public's general aversion to it. Yet, in the long term, they might prove to be efficient and even fair. For that reason, if the process would be rendered fully transparent, politicians might interfere in the process for political reasons and even stall it at an early stage, rather than after reflecting over its overall implications.[176] Full and immediate transparency to the entire public of the overall "usage" process might lead to an outcome which might be counter to the public's long-term interests.

[173] Schauer, *supra* note 18, at 1348–50.

[174] *See* Taipale, *supra* note 80, at 17, 39–48 (discussing the process of blocking the TIA project, and how it is probably ongoing); Cate, *supra* note 13, at 450–51; *see also* CRS REPORT, *supra* note 62, at 7–8.

[175] For this argument in general, see Schauer, *supra* note 18, at 1353.

[176] According to one commentator, the demise of the TIA is a clear example of this dynamic. Kim Taipale argues that the TIA was a balanced initiative which includes sufficient privacy safeguard, and its cancellation and subsequent shift to other more secretive realms did more harm than good. *See* Taipale, *supra* note 80, at 49.

To conclude, our discussion here demonstrates that both the "shaming" and "market (or political) forces" arguments have limited force in promoting transparency to the entire public. These arguments are convincing when they pertain to decisions made by high-ranking officials and are clearly contrary to social norms that are broadly held. The arguments for such transparency also have merit when the practices they reveal are understandable, or at least easily built into a convincing narrative. In all other contexts, this transparency theory might be unable to justify the costs and detriments it generates. Interestingly, the current legal landscape described above seems to account for this distinction. It mandates disclosure of the broad, central decisions governing the predictive process. It leaves the somewhat technical details of ongoing operations outside public view. It also leaves the trends of actual use opaque, an outcome which could be justified given the political tendency to react to the short-term yet possibly misplaced concerns which these disclosures might raise.

2. *Transparency and Crowdsourcing*

Transparency might enhance the accuracy and fairness of predictive models in a very different way. Rather than incentivizing effective governmental actions, transparency can incorporate knowledge from *outside* the government into these processes so as to improve the final outcome. Generally, the level of expertise, time, and attention available outside the specific agency (and even government in general) are greater than the knowledge available within.[177] Exposing more information regarding the inner works of government to a broader segment of the public will enhance the chances to receive meaningful feedback.[178]

These arguments are closely linked to another facet of recent legal scholarship which addresses peer production: the mass participation of individuals from varied walks of life and with different skill sets in joint projects. As Yochai Benkler[179] and others[180] explain, IT and especially the Internet led to the rise of a third collective/industrial force which matches and even surpasses that of the firm and the market. Transparency can enable these powerful dynamics and promote governmental objectives. Indeed, scholars have called for transparency to facilitate this dynamic of participation in governmental projects in a

[177] For various sources regarding this argument, see Shkabatur, *supra* note 28, at 1450.

[178] Fenster, *supra* note 152, at 100–02.

[179] YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 59 (2006), *available at* http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf.

[180] A great deal of popular writing has flourished in these fields. *See e.g.*, CLAY SHRIKY, HERE COMES EVERYBODY: THE POWER OF ORGANIZING WITHOUT ORGANIZATIONS 55, 109 (2008); JAMES SUROWIECKI, THE WISDOM OF CROWDS 85 (2004); DON TAPSCOTT & ANTHONY D. WILLIAMS, WIKINOMICS: HOW MASS COLLABORATION CHANGES EVERYTHING 268 (2006).

variety of similar contexts. For instance, in the context of system security, Peter Swire argues that the government can assure and enhance its systems by providing the public with the code for its software.[181] Others argue for expanding transparency of automated governmental decision making[182] for similar reasons.[183]

Crowdsourcing will prove effective only if transparency will lead to significant feedback from outside the government. This point indicates the limits of this transparency-enhancing theory. Whether such external feedback will indeed transpire in the specific context of predictive modeling is a difficult question. The key to the answer is probably whether there is significant knowledge outside the government regarding the relevant issue and motivation to convey it.

In the context of data mining and prediction, knowledge surely exists. Academia and the private sector are leading the way in these fields.[184] The knowledge crowdsourcing can bring into the governmental realm from external sources pertains to almost all stages of the prediction process. Experts and layman from a variety of disciplines can provide meaningful insights regarding methods of aggregating data and engaging in data-mining analyses. They can also examine theories of causation and assess the feedback provided at the "usage" stage. Above all, experts can work through the code of the software operating these schemes, examining its neutrality and whether it indeed achieves the tasks it purports to carry out. Therefore, while this theory can apply to all the process's segments, it is usually linked to the "analysis" stage (designated B. in Table 1, *supra*) of the predictive process. Here, the disparity between governmental knowledge and relevant, yet freely available, expertise appears to be the greatest.

It is also fair to assume that there is sufficient motivation for participating in the crowdsourcing dynamics at this juncture. Some of the motivations transpiring in other contexts (such as the open source movement, which included a powerful adverse response to the Microsoft monopoly) will not play out in the governmental automated prediction context.[185] Several other incentives are extremely relevant, however.

[181] Peter P. Swire, *A Model for When Disclosure Helps Security: What Is Different About Computer and Network Security?*, 3 J. TELECOMM. & HIGH TECH L. 163, 168–69 (2004). Swire argues that the benefits of contributing "white hat" hackers will surpass the risks of easily showing the systems shortcomings to the "black hat" ones.

[182] Citron, *Technological Due Process*, *supra* note 30, at 1308–13.

[183] Citron, *Open Code Governance*, *supra* note 33, at 366. For a critical view of how these dynamics are applied to the governmental and municipal context, see Shkabatur, *supra* note 28, at 1419, 1436, 1443.

[184] *See* Shkabatur, *supra* note 28, at 1450 (providing references for other technology-related contexts).

[185] In other contexts (such as open software source and content projects) scholars indicate that individuals might be motivated by spite to develop products competing to those of Microsoft. In yet other contexts, spite might prove to be a motivator to "get back" at a bad employee or vendor, and in that way inform the public of their wrongdoings. *See* Shmuel I. Becher & Tal Z. Zarsky, *E-Contract*

Individuals will contribute to this project altruistically.[186] Others will do so as a hobby, pastime, or as a part of their academic research. Some might do so as means to contribute to a community which might be emerging[187] or to enhance their reputation within their social or professional circle.[188] The success of recent technology-related crowdsourcing schemes carried out by local governments strengthens these assertions.[189]

Current transparency regulation regarding automated prediction does not reflect any aspects of this theory. The most practical segment for implementing transparency in view of this theory is where its absence is most noticeable—regarding the computer code charged with running the analysis. Rather than allowing experts to review and comment on it, the government provides almost no insights as to the code's inner workings.

The most obvious challenge to the "crowdsourcing" justification is that the level of transparency it requires generates great vulnerabilities. Allowing external entities to influence (and possibly manipulate) policy on such delicate matters is unthinkable to many bureaucrats. Even scholars addressing crowdsourcing strategies concede that national security might not be a fitting context for their implementation.[190] An additional problem is that these forms of disclosure might allow for the leak of trade secrets from contractors developing software for the government.[191]

While translating this rationale into transparency policy is important, it could be substituted at times by providing information to a selected group of institutions and experts. These experts will assist the relevant government agency with feedback on predictive modeling projects without disclosing the information further. While benefits might be lost by such narrow exposure, these will not be vast; recent

Doctrine 2.0: Standard Form Contracting in the Age of Online User Participation, 14 MICH. TELE. & TECH. L. REV. 303, 336 n.145 (2008).

[186] For a different perspective, which plays down the current level of contribution to open source projects which is altruistically motivated, see generally Jonathan M. Barnett, *The Host's Dilemma: Strategic Forfeiture in Platform Markets for Informational Goods*, 124 HARV. L. REV. 1861, 1874 (2011).

[187] *See* Citron, *Open Code Governance*, *supra* note 33, at 383–84 (stating that interested parties of skill will participate in these projects, mentioning reputation as a driver of such actions and participation).

[188] BENKLER, *supra* note 179, at 79; ERIC RAYMOND, THE CATHEDRAL AND THE BAZAAR: MUSING ON LINUX AND OPEN SOURCE BY ACCIDENTAL REVOLUTIONARY (1999); Lior Jacob Strahilevitz, *"How's My Driving?" for Everyone (and Everything?)*, 81 N.Y.U. L. REV. 1699, 1701 (2006).

[189] Shkabatur, *supra* note 28, at 1450, 1455.

[190] Citron, *Open Code Governance*, *supra* note 33, at 384–87. In addition, Swire, *supra* note 181, at 186, explains that the security realm includes specific attributes that make data exposure less severe. Namely, he explains that in many contexts vulnerabilities are apparent to any adversary. *Id.* at 194. It is interesting to note that the same could be said of some contexts of this discussion as well.

[191] Citron, *Technological Due Process*, *supra* note 30, at 1292–93.

experiences with governmental crowdsourcing experiments indicated contributions were only limited to a concentrated pool of experts.[192] If these could be located in advance, narrowing the scope of transparency would have a limited negative effect.

To conclude, when structuring disclosure recommendations premised upon the crowdsourcing justification, a balance must be struck between sharing the information with the entire public and specific institutions. It is also interesting to contrast this theory against the previous one mentioned (which focused on shaming and political forces). In the previous discussion, the theory's weakest points were justifying transparency for technical matters and decisions, but such issues are at the core of the "crowdsourcing"-based justification for transparency.

3. *Transparency, Autonomy, Privacy, and "Notice"*

In a predictive modeling scheme, which is premised upon the analysis of personal information, calls for transparency flow directly from an additional theoretical premise—the information privacy rights of individuals whose personal information was used throughout the process—the "data subjects."

The basic premise leading to this aspect of transparency relates to the notion of control individuals have over their personal information.[193] This theoretical notion has been broadly accepted in the European Union (EU),[194] as part of data protection law. It was only partially recognized in the United States.[195] This notion of "control" could be understood as an extension of the individual's autonomy and ability to control his personal information.

The idea of control over personal information has made its way from theory to practice. It was translated into several concrete principles that after several transitions[196] formulated the "Fair Information Practices" (FIPs).[197] FIPs are gaining force as a universal standard of

[192] Shkabatur, *supra* note 28, at 1453, 1462.

[193] ALAN F. WESTIN, PRIVACY AND FREEDOM 23 (1968). For a critique of this theory, see Paul M. Schwartz, *Beyond Lessig's* Code *for Internet Privacy: Cyberspace, Privacy Control, and Fair Information Practices*, 2000 WIS. L. REV. 743, 746. This theory was promoted in the technology context of information privacy by Lessig, see LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE 143 (1999).

[194] Council Directive, art. 10, 1995 O.J. (L 281) 41 (EC) [hereinafter *EU Data Protection Directive*]; DANIEL J. SOLOVE ET AL., INFORMATION PRIVACY LAW 38 (2d ed. 2006).

[195] See 5 U.S.C. § 552a(d) (2006). *See* also *infra* Part IV.B.4.

[196] For a recent discussion of the formulation of Fair Information Practices (FIPs), see ROBERT GELLMAN, FAIR INFORMATION PRACTICES: A BASIC HISTORY (2012), *available at* http://bobgellman.com/rg-docs/rg-FIPShistory.pdf.

[197] *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (Sept. 23, 1980), *available at* http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

privacy and data protection rights. In the United States, FIPs were implemented in several laws,[198] including the Privacy Act.

The notion of control over personal information pertaining to an individual is featured in FIPs through the principle of "choice."[199] Choice constitutes the ability to block future uses of personal information, unless the data subject provides specific consent to the relevant form of use. In the context of governmental prediction, however, the information is usually collected by government without the individual's consent.[200] In addition, in these settings there is very limited control regarding subsequent uses.[201]

The inability to fulfill the obligations of these most basic elements of FIPs brings us back to transparency. Beyond "choice," other privacy- and autonomy-related principles could still be fulfilled, such as the principle of "openness" or "transparency." These principles were central to FIPs from their earliest stage.[202] In FIPs' current version, this notion is encapsulated in the principle of "notice." Notice commonly refers to informing individuals that personal information about them is being collected and used.[203] Thus, recognizing the importance of control over personal information, as well as the principle of notice—which is perhaps the only limited way in which control could be afforded in the governmental prediction context—should lead governments to engage in greater transparency.

To some extent, this theory is anchored in existing law. As mentioned, notice is reflected in the Privacy Act.[204] The Act provides individuals with some information regarding the collection and use of their personal information even when part of an automated predictive process. This requirement is met through the publishing of SORNs, available on governmental websites.[205]

In practice, however, this application of the notice right does not significantly promote transparency. The information conveyed through SORNs or other public disclosures is limited, general, and rarely accessed or understood by the public (be they data subjects or not). Furthermore, the information provided only pertains to the "collection"

[198] *See* Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get)*, 2001 STAN. TECH. L. REV. 1, at 14.

[199] *Id.* at 3.

[200] *See, e.g.*, ATS-P PIA, *supra* note 118, at 20 (explaining how information is usually collected by the government through various interactions which cannot be avoided (such as crossing a border); thus, there is no practical, real option to decline collection).

[201] *Id.* at 21 (indicating very limited control over such restrictions).

[202] See GELLMAN, *supra* note 196, at 4, 10; *see also* Joel R. Reidenberg, *Setting Standards for Fair Information Practice in the U.S. Private Sector*, 80 IOWA L. REV. 497, 515 (1995).

[203] Reidenberg, *supra* note 202, at 515.

[204] *See* 5 U.S.C. § 552a (2006).

[205] *Id.* at § 552a (e)(4). *See System of Records Notices (SORNs)*, U.S. DEP'T. OF HOMELAND SECURITY, http://www.dhs.gov/system-records-notices-sorns (last visited May 20, 2013).

stage (designated A. in Table 1, *supra*). This disclosure process does not provide insights as to the way the analysis is conducted and how its outcomes are applied. Even within the collection stage, notice is not always perceived to include the right to understand how the database is formulated, or how similar fields are aggregated into one dataset which is later used. Yet notice and the underlying notion of control might have additional implications for our discussion of transparency. We, therefore, must further examine whether this theory calls for broader transparency norms for automated prediction schemes in two dimensions—the various stages of the process and those who should be provided with access to the information.

Let us first examine the data recipients. On its face, this theory can only justify disclosure to those affected by the process—the data subjects. A strong argument could be made to expand the right to receive information regarding the prediction process to the entire public. As the prediction methods require the analysis of personal information, which pertains to almost the entire population, disclosure, according to this theory, also must be provided to all as well. Everyone is a "data subject" one way or the other.[206] To a certain extent (and as noted above) this is the actual governmental practice; the SORNs addressing the use of information are made available to the entire public. In other words, this privacy and autonomy based argument can be transformed into an overall transparency right when mass data analysis takes place.

Next, we must inquire whether the "notice" principle, which is derived from the notion of data subject autonomy and control, can justify broadening transparency into the latter stages of the prediction process beyond the initial collection stage noted above. Yet reaching farther into the analysis process is probably a theoretical stretch. On its face, one can argue that the latter steps of the data flow (the "analysis" and "usage" stages) all result from the initial secondary uses of personal information. The data subjects' autonomy and liberty should be acknowledged by providing them with a full view of subsequent information flow, all resulting from their control over personal information pertaining to them.

Yet such arguments for the expansion of this theoretical justification to the latter steps of the process should mostly be rejected. Doctrinally, arguments regarding "control" over personal information have limited force in U.S. legal thought.[207] Unlike the European Union, the United States did not adopt a universal recognition of the right of data protection. In the commercial realm, individuals currently have

[206] Obviously, this argument is only plausible in instances where providing such information could be done without harming the privacy rights of other data subjects.

[207] Many have argued that this should change. For a recent and balanced discussion which also takes into account the possible "property-like" attribute of privacy protection, see Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2056, 2060, 2076 (2004).

almost no control over subsequent data uses.[208] Thus, it is problematic to rely upon this notion of "control" to develop an elaborate and far-reaching transparency scheme.

This conclusion follows from a normative analysis as well, which goes beyond the confines of this Article and therefore will be noted very briefly. The "privacy as control" theory (the premise of the entire analysis set forth in this segment) is losing ground due to technological innovation and social changes. This theory is premised upon the notion that individuals value their autonomy as it is reflected in their control over their personal information. This notion is constantly challenged as individuals quickly concede control over their personal information in a variety of commercial and social settings—showing that they either see no need for such control, or attach a low value to it. In response, one can argue here that for various reasons, individuals are unable to correctly establish, at least at first, the proper value of such control. Therefore, an alternative theory will note that individuals must be provided with control so that they may derive many social benefits, even if they do not intuitively understand them at first.

This alternative theory and regulatory response has limits as well; it calls for a paternalistic intervention in which the state signals to its citizens which values they must attach to their personal information. In addition, given the complexities of information flows, providing extensive control over personal information to individuals encumbers the analysis process and in that way might limit some of the benefits derived from subsequently analyzing such personal data. For these reasons, even if the "privacy as control" theory is to be accepted in some contexts, it must be applied narrowly, while assuring the protection afforded is not overreaching, and that regulators' and policymakers' moral judgments are not substituted for their constituents' preferences.

It is therefore challenging to apply the "control" theory to justify the individual's control over the latter steps of the information flow process. Indeed, "control" is usually linked to an understanding as to how the information is collected and perhaps aggregated. Yet it would be difficult to derive from this theory the ability to control subsequent analyses and even uses of the information—uses that perhaps only impact other individuals and do not affect the "data subject." Furthermore, this theory cannot justify some of the more specific forms of transparency in the latter stages of the process. The above analysis indicates the need for ancillary information rights to assure that the disclosure is effective. For instance, transparency could potentially be understood as mandating governmental studies of the causation

[208] For an exception, see The Cable TV Privacy Act of 1984, 47 U.S.C. § 551 (2006). *See also* SOLOVE ET AL., *supra* note 194, at 191. For an overall mapping of the distinction between the EU and the US regarding this matter, see PETER P. SWIRE & ROBERT E. LITAN, NONE OF YOUR BUSINESS: WORLD DATA FLOWS, ELECTRONIC COMMERCE, AND THE EUROPEAN PRIVACY DIRECTIVE (1998).

underlying its actions, and the success and failure rates of the project. The connection between such information and the control rights of data subjects is extremely weak. The theory of autonomy and control is not sufficiently robust to cover these elements.

To conclude, this privacy-, autonomy- and control-based theory has only limited relevance to the issue at hand. It can surely promote transparency at the early stages of the process. It can justify the process of providing data subjects with information about the data collection. If considerably stretched, it could justify providing information about the forms of data analysis (which can still be understood to rely on the personal data of individuals, and thus affect their autonomy). But the subsequent steps of the process are beyond its analytic reach. Acknowledging this theory in the automated prediction context is important, however. It promotes the understanding that transparency policy in this unique context must also account for an additional set of concerns—those pertaining to information privacy, and control over personal information—all related to the overall notion of personal autonomy.

4. Autonomy of the Impacted Individual

Autonomy-based theories generate justifications for transparency from a very different perspective—that of the individuals adversely affected by the outcomes of the predictive process. These are the individuals held up at the airport or engaged with an extensive IRS audit. If individuals are affected by predictive modeling, they presumably have a right to understand why. They should receive an explanation as to the decision criteria and to the logic behind these actions. In other words, these affected individuals deserve some form of transparency.

This intuitive transparency notion can easily be framed in terms of autonomy.[209] An individual has a right to learn the reasons for events which affect her.[210] Such information empowers her, and she senses she is treated with respect, as a human being. This notion is deeply embedded in European law and specific member states.[211] In the United States,

[209] This abstract notion is articulated by Cohen as follows: "We would not tolerate comparable restrictions on access to the basic laws of phyics, chemistry, or biology, which governs the operation of the physical environment." COHEN, *supra* note 12, at 235.

[210] This right is closely related to the European understanding of privacy and data protection as a notion closely related to metaphors stemming from the works of Franz Kafka; *see* Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393 (2001); MARY DEROSA, DATA MINING AND DATA ANALYSIS FOR COUNTERTERRORISM 14, 16 (2004), *available at* http://csis.org/files/media/csis/pubs/040301_data_mining_report.pdf.

[211] In the Netherlands, Staatsblad van het Koninkrijk der Nederlanden [Stb.] art. 42(4) (Neth.). In Spain, Organic Law on the Protection of Personal Data art. 13(3) (1999, 23750) (Spain). In the EU, Council Regulation 45/2001, art. 13(d), 2000 O.J. (L 8) 10 (EC) (applying the directive to EU government bodies). Note, however, that these rules include exceptions for security and law enforcement.

beyond autonomy, scholars strived to embed these specific concerns in the notions of the U.S. Constitution, while referring to the concepts of Due Process and the protection provided by the Fourth Amendment.[212]

This last statement was purposefully carefully worded. The analysis here addresses mere constitutional notions, as opposed to doctrinal principles which will provide legal protection. Predictive modeling, as described here, does not fall within the confines of actions that mandate actual constitutional protection. It usually does not trigger "due process" protection. Due process calls for a relatively high threshold of harm—to life, liberty, or property.[213] The harms associated with actions addressed in this Article (and those mentioned in the examples above)—additional costs, time, aggravation, and some financial burdens—do not fall within such a definition of harm.[214] In addition, the doctrine of "criminal process preemption" leads to setting "due process" protection aside in all instances involving criminal investigations (searches, surveillance, stops)—the instances discussed in this Article.[215]

The Fourth Amendment cannot be directly invoked in these situations either. The process here described addresses three actions which might be considered relevant to Fourth Amendment scrutiny: the data collection, analysis, and subsequent use. Yet they all fall outside its protection given the Article's underlying basic assumptions—that all information was legally collected by government or commercial entities.[216] In that case, subsequent analyses of the data are rarely prohibited by the Fourth Amendment.[217] Finally, the Article addresses cases in which the predictive analysis itself did not lead, on its own, to any action which implicates constitutional protection. Therefore, predictions used for airport examinations and financial audits are outside the realm of the Fourth Amendment.[218]

[212] . U.S. CONST. amends. IV, V, XIV.

[213]. Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 23 (2005).

[214]. *Id*. at 48–49.

[215]. *Id*. at 23–25.

[216]. The underlying premise of the assumption in the text is that individuals have no expectation of privacy vis-à-vis government (in terms of Fourth Amendment protection) if the information was consensually provided to (even commercial) third parties. Smith v. Maryland, 442 U. S. 735, 743–44 (1979). This notion is known as the "third party doctrine," which is commonly critiqued by legal scholars. *See*, *e.g.*, Katherine J. Strandburg, *Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change*, 70 MD. L. REV. 614 (2011). The days of this doctrine might be numbered, however. Recently, in an important concurring opinion, Justice Sotomayor questioned the wisdom of this rule, by noting how she "*would not assume that all information voluntarily disclosed to some member of the public for a limited purpose is, for that reason alone, disentitled to Fourth Amendment protection.*" United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) (emphasis added). If the doctrine is to change, then obviously the statement made in the text must be revised as well.

[217]. *See* Cate, *supra* note 13, at 460–61; Slobogin, *supra* note 11, at 329.

[218]. *See* Steinbock, *supra* note 213, at 26.

All that being said, Due Process Clause principles can provide a normative guideline for future legislators approaching these matters. Indeed, in an insightful article touching upon some of these issues, Daniel Steinbock examines the role of due process in prediction processes automated by data mining.[219] He finds that measures which resemble "due process" rights should be set in this context, even though the constitutional protection does not apply.[220] These measures are appropriate in view of the individual's dignitary rights, which might be comprised through this process. He states that these rights should include some form of notification of the process the individuals were subjected to—or, in other words, some form of transparency in the prediction process.[221]

Transparency rights might follow from Fourth Amendment principles as well. While the theories and rationales behind Fourth Amendment protection introduce a variety of aspects, transparency could be tied into this important context by recognizing a hidden concern the Fourth Amendment strives to mitigate—*targeting harms*. In this context, Sherry Colb's work on the theory of the Fourth Amendment is perhaps the most illuminating. Colb states that the Fourth Amendment should also be understood to include rights to mitigate "targeting harms"—the sense that an individual is being "targeted" by the government for no proper and acceptable reason.[222] Colb explains that targeting harms could be mitigated by ancillary rights, such as the right to understand that the reason for being selected (or "targeted") was not arbitrary.[223] In other words, she refers to the right to know there *is* a rationale behind the selection process. Recognizing that such a right could be derived from the Fourth Amendment brings us back to the notion of transparency. Providing transparency in the predictive modeling process can indeed counter such targeting harms, by ensuring those affected by the process were not subjected to an arbitrary selection process. To conclude, both arguments premised on Due Process and Fourth Amendment interests call for providing transparency to a specific segment of the population—those impacted by the prediction process.

While recognizing the justification for transparency is relatively convincing and straightforward in general, applying it in practice raises difficult questions. European laws[224] and U.S. scholarship addressing transparency requirements, which indeed reflect this rationale, call for

[219] *Id.* at 7.

[220] *Id.* at 81–84.

[221] *Id.* at 65–70.

[222] Sherry F. Colb, *Innocence, Privacy, and Targeting in Fourth Amendment Jurisprudence*, 96 COLUM. L. REV. 1456, 1464 (1996).

[223] *Id.* at 1489–93 (emphasizing the subject's perception of the targeting practices).

[224] *See supra* note 211.

informing the affected individuals of the profile they were subjected to.[225] In terms of this Article's framework, it provides (or calls for providing) information regarding the "usage" stage (designated as C. in Table 1, *supra*) while focusing on the actual factors selected for prediction. With such information in hand, the affected individuals would be able to object if they find it to be inaccurate.

These requirements, however, were shaped with the most basic profiling practices in mind. In the age of data mining, conveying information regarding the profile used might be either meaningless or impossible. The "profile" might prove to be a string of parameters that indicate a problematic correlation the affected individual falls within. When provided with such information, the individual could rest assured there was no identification error. She might further understand that the process was not completely random. Yet without understanding the inner workings leading to this outcome, these results might still appear arbitrary and wrong. Therefore, this limited form of transparency might not be sufficient to restore autonomy and dignity with regard to the automated prediction process.

To empower relevant individuals and provide meaningful feedback, the theoretical justification noted calls for expanding the notion of transparency within the "usage" stage. It requires that the automated prediction process is interpretable—namely, that the selection process is explainable to humans should the need arise. Here, one could even further argue that dignity interests require that causation and not mere correlation is found prior to launching action.[226] With such additional disclosure, individuals can obtain sufficient insight to the process and how it relates to their lives.

Furthermore, this latter transparency justification calls for other forms of transparency—enhancing steps at the "usage" stage—especially those described above as providing an *ex post* perspective. To assure dignity and lack of targeting, the individual should receive assurances as to the precision, effectiveness, and lack of discrimination in the process. The information provided through the feedback studies mentioned above can promote these objectives.

As mentioned above, we must also examine whether this theory could promote transparency in other stages of the prediction process. On its face, this autonomy-based theory could justify transparency at the "analysis" stage (designated as B. in Table 1, *supra*) as well. Information regarding the specific prediction process applied is essential for allowing the affected individual to retain autonomy and dignity. With information

[225] Citron, *Technological Due Process*, *supra* note 30, at 1305.

[226] The issue of correlation vs. causation at this juncture raises fascinating questions that cannot be properly addressed within the confines of this Article. I hope to further address them in future work.

regarding these important steps of the process, the mere correlations used (in which an individual was implicated) can be understood as part of a broader picture. This will prove helpful in understanding that targeting was not arbitrary, and perhaps even in challenging its findings. A case for further expansion of transparency on the basis of this theory to earlier stages of the process (i.e., the "collection" stage), cannot be made convincingly, however.

Thus far, this discussion has set forth a powerful argument which promotes transparency in a substantial part of the predictive modeling process. The theory as presented thus far, however, has severe limitations as it only pertains to a small segment of the population—those adversely impacted by the relevant predictive practices. After further consideration, however, it is possible that this theory could endorse transparency of the broadest form—to the entire public.

To make this point, it is first important to concede that disclosures made to impacted individuals might quickly make their way to the broader public anyway. Those adversely affected will provide their information online (even if stigma may attach, they might do so anonymously). With time, the pieces of the puzzle will come together, and a full picture of the overall predictive practices would emerge in the public realm. With this insight in mind, the transparency justification discussed here can endorse disclosing the information addressed above to the entire public, rather than parceling it out to a small segment of society. After all, if the government reveals this information to impacted individuals, it might as well initially provide it to the public, and in that way, also achieve the broader transparency objectives noted throughout this Article. Indeed, engaging in selective disclosure is not a costless endeavor and limits other social benefits of transparency.

Yet this last call for full disclosure is of limited force. To further strengthen this argument, we must also show that society has much to gain—and government little to lose—from expanding disclosure in this way. Let us move to address these two points. Expanding disclosure carries social benefits. It is quite possible that with limited disclosure, which does not go beyond the mere "affected individuals," the dynamic of online knowledge sharing will lead to a partial and, in many instances, biased and wrong overall picture of governmental practices. With partial disclosure, the information flowing to the general public might reflect, for instance, that the government relies on factors that are racial and discriminatory by nature (while in fact it is not!). Inaccurate (or even biased) information regarding automated predictive practices can have devastating outcomes. Individuals will change their behavior to avoid additional scrutiny. Segments of the population will falsely believe they are targeted. Therefore, it is within the government's interest to fully reveal its strategies to the public. The information would be leaked anyway, so the government can at least assure that the data made available is correct.

Government has little to lose from such enhanced exposures to the entire public if information is available to the affected individuals. This argument might be unpopular at first. Those arguing for the benefits of governmental opacity (to be discussed below) will state that even with limited mandatory disclosure in place, a certain level of opacity regarding governmental practices in the context of automated predictions could be maintained. Yet transparency according to this theory, when accepted, places information regarding the inner actions of government in the hands of those indicated as higher risks. It would be quite difficult to argue that transparency to the entire population would harm government interests if information was already provided to the affected individuals.

To conclude, this final theory provides powerful arguments for providing transparency regarding the "back end" of the prediction process—the "usage" and "analysis" stage. It sets forth strong theoretical reasons for disclosure to a limited slice of the public ("affected individuals"), which could be broadened to the entire public given practical technological developments.

5. Unconvincing Transparency Theories: Automation and Autonomy

The practices of predictive modeling feature several unique traits. One of them is the extent of automation embedded in the process. While the process is not fully automated and includes a substantial role for human analysts, the role of computerized analysis and decision making goes beyond existing practices. Arguably, an automated process impedes the impacted individual's autonomy by its nature. The cure to such impediment is to provide additional insights regarding the process. In other words, the mere fact that the process features a high degree of automation calls for enhanced transparency. This theoretical argument has been set forth and even legislated in the EU. The argument, however, is unconvincing and therefore should not have an impact on transparency policy in the U.S.

In the EU, specific legal rules addressing automated processes are set forth on several layers.[227] The Data Protection Directive addresses this issue directly in Article 15,[228] with a supplemental right in Article

[227] For an excellent recent discussion of this issue, see Douwe Korff, *Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments* (European Comm'n Directorate-Gen. Justice, Freedom and Sec., Working Paper No. 2, 2010), *available at* http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf.

[228] *EU Data Protection Directive*, *supra* note 194, at art. 15. Similar rights were provided in a recent EU draft to update the existing data protection directive. *Commission Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals With Regard*

12(a)[229] to receive information about the underlying logic of the process. Additional laws were set forth on the state level as well[230] and vary in the contexts in which they apply.[231] They usually concern important administrative decisions impacting individuals. The redress this provision affords also varies among EU members. Some states provide a right to stop the process completely, while others call for an individual (human) supervisor to review the case. The most popular remedy, however, is providing individuals with the "logic" behind the automated process.[232] Using the terminology set forth in this Article, these laws provide transparency to the affected individuals regarding the "usage" stage (designated as C. in Table 1, *supra*) of the process.

Some evidence shows that even within the EU, many are unconvinced of the necessity and logic of these transparency-for-automation rules. A recent report summarizing data protection law in the European Union tells an interesting story when examining this issue from a "law in action" perspective. The report indicates that almost no cases are brought regarding these issues to European courts—a surprising fact as this right is perceived to be commonly violated.[233] It further notes that the states' Data Protection Authorities (DPAs) show a very low level of activity concerning this matter.[234] The report finally mentions the limited academic writing on automation and the relevant section of the Data Protection Directive.[235] It therefore concludes that these obligations should be urgently clarified, as many future actions will fall within this provision.[236]

There is also no clear indication as to the philosophical foundation of these requirements in Europe. It is fair to assume that they are premised upon broad notions of individual autonomy, and more

to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), COM (2012) 11 final (Jan. 25, 2012).

[229] *EU Data Protection Directive*, *supra* note 194, at art. 12(a).

[230] Korff, *supra* note 227, 82–86.

[231] Note that the EU Data Protection Directive does not pertain to law enforcement activities. *Supra* note 194. These are governed by a different set of data and EU rules. A recent EU framework addressing this issue is Council Framework Decision 2008/977/JHA. Article 7 of this Framework notes that automated searches require safeguards from relevant governments. "*A decision which produces an adverse legal effect for the data subject or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to the data subject shall be permitted only if authorised by a law which also lays down measures to safeguard the data subject's legitimate interests.*" Council Framework Decision 200819771JHA, art. 7, 2008 O.J. (L 350) (EC) (emphasis added).

[232] Korff, *supra* note 227, at 82–86.

[233] *Id*. at 98.

[234] The British privacy authority has gone as far as to substantially limit its interpretation of these provisions, in terms of the entities it pertains to and the rights it provides. *Id*. at 85.

[235] *Id.* at 85, 119 n.138.

[236] *Id*. at 23–24.

specifically the notions of dignity and decency.[237] They derive from the understanding that individuals should be treated as fellow persons and not mere machines. The European aversion and suspicion towards automation might also be derived from neo-Luddite beliefs regarding the inferiority of computer searches.[238]

The theories calling for the limitation and specific treatment of automated processes are unconvincing. Linking the lack of dignity and automation is, I believe, an anachronistic notion. In the twenty-first century, one need not fear a computerized process. If computerized searches can provide fair and efficient outcomes, should they still be considered as undignified? Indeed, safeguards (through either transparency or other measures) should be applied to all steps which might compromise rights of individuals and seem arbitrary, be they automated or manual. The level of automation needs not, on its own, merit a higher level of transparency.[239]

Additional theories that justify cautious treatment (which includes enhanced transparency) toward automated processes should be rejected as well. First, some argue that these dynamics generate errors and thus mandate greater disclosure.[240] This might be true, yet the extent of errors in automated processes should not be considered alone, without addressing other factors. Alternative strategies (to the automated prediction process) lead to errors as well,[241] and human decisions carry particular risks of their own—such as hidden and internal biases that might be premised upon bigotry.[242] Therefore, the existence of errors in automated processes is not a convincing factor for categorically mandating extensive disclosure in predictive processes.

Another argument states that automatic decisions are less likely to be doubted. Computerized automations generate an (erroneous) aura of flawless decision-making abilities.[243] This is indeed a serious concern. I doubt, however, whether additional transparency provides a sufficient answer. This concern could be resolved through other measures such as educating the public and relevant decision makers of the true nature of

[237] General Data Protection Regulation, *supra* note 228, at art. 1. Korff finds that this rule originated from French law, where indeed it was premised upon principles of dignity. Korff, *supra* note 227, at 82. Lee A. Bygrave, *Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMPUTER L. & SECURITY REP. 17, 21 (2001), *available at* http://folk.uio.no/lee/oldpage/articles/Minding_machine.pdf (noting the threat to human dignity created by automation).

[238] Cohen provides a recent, interesting view of the Luddite movement. *See* COHEN, *supra* note 12, at 270. Cohen explains that this movement set forth an important social alternative, rather than mere conservatism. *Id.*

[239] For more on this argument, see Tal Z. Zarsky, *Governmental Data Mining and its Alternatives*, 116 PENN. ST. L. REV. 285 (2012).

[240] Citron, *Technological Due Process*, *supra* note 30, at 1278.

[241] *See* Zarsky, *supra* note 239, at 299.

[242] *See* Zarsky, *supra* note 12, at 22.

[243] Bamberger, *supra* note 21, at 675.

automation.[244] To summarize, in today's age of information glut, some form of automated analysis is essential. Automation generates specific problems, which require articulation. The four additional theories set forth above: fairness, efficiency, crowdsourcing, and autonomy (both in term of data control of the subject and understanding of the impacted individual) provide such a response.

V. TRANSPARENCY—WHY NOT?

While scholars and policymakers argue forcefully for transparency, powerful counterarguments are set in place at almost every juncture. Some see transparency as a disease causing effective and active government to die.[245] Others point to "enormous unintended consequences" which might flow from transparency requirements, and warn that these might generate great detriments.[246] This Part discusses central arguments against transparency claims that pertain to the process of automated prediction, which note that (1) transparency undermines governmental objectives, and (2) it generates concerns related to stigma and stereotyping.

Before proceeding, a short note on *costs* is called for. Facilitating transparency generates costs on various levels. Transparency calls for writing up, editing, collecting, and disseminating information.[247] It also calls for additional active research. The costs of these actions could be substantial. Given the importance of transparency to the democratic discourse as well as personal autonomy, these costs are a price society must bear, if transparency is indeed called for. Thus, for the rest of this discussion, we set arguments concerning costs aside.[248]

A. Undermining Governmental Objectives

1. Transparency and Sidestepping Proxies

The most common objection to transparency states that it would undermine the objective the predictive model strives to achieve. The IRS's prediction models strive to advance compliance with the tax code and increase tax revenue. Transparency will arguably enable tax

[244] Citron, *Technological Due Process*, *supra* note 30, at 1305.

[245] Stuntz, *supra* note 148.

[246] Fenster, *supra* note 152, at 906 (citing James T. O'Reilly, *FOIA and Fighting Terror: The Elusive Nexus Between Public Access and Terrorist Attack*, 64 LA. L. REV. 809, 812–14 (2004); Scalia, *supra* note 164, at 15).

[247] Fenster, *supra* note 152, at 937. Samaha notes the costs are about $320 million per year for FOIA and Privacy Act requests, but continues to note they are "not staggering." Samaha, *supra* note 150, at 959.

[248] Note that opacity has costs as well; the government spends a great deal of money classifying documents, and thereafter securing them. Fenster, *supra* note 152, at 900.

avoidance and lead to lower tax revenue. The same could be stated regarding law enforcement and national security prediction projects, yet allowing adversaries to sidestep the predictive model would lead to far graver consequences.[249] This powerful rationale is reflected in current law. Every disclosure law has a law-enforcement exemption clause.[250]

This argument has a very powerful intuitive appeal. Yet a closer look shows that it cannot be equally applied to all forms of disclosure, regardless of the process's stage. Therefore, this general notion must be broken down into several underlying rationales. The simplest way to understand this argument is that knowledge of the inner workings of the automated prediction models in the hands of adversaries will allow them to "game the system."[251] Given the automated nature, they would be able to do so systematically, with limited risk of getting caught. A computer—so we might intuitively believe—sticks with the program in a way that is itself predictable. Therefore, automated processes are most susceptible to these forms of "attacks."[252]

This problem transparency generates pertains to several of the stages within the prediction process, as drawn out above. For instance, it pertains to the "collection" state (designated as A. in Table 1, *supra*). Information about the datasets used in the process could assist evildoers. With such information in hand, they would know which datasets to avoid falling into, and in that way remain "off the grid" and outside the realm of government analysis and detection. These forms of disclosure, however, will have limited value to evildoers (and thus their exposure to the public is of limited detrimental value) if the realm of databases collected and used are so vast that sidestepping the governmental initiative would be nearly impossible—e.g., consider a list of all credit card issuers providing information to the government. In view of the vastness of governmental initiatives, this indeed might be the case and thus these arguments for opacity are without basis.

Note, however, that the "collection" stage includes other forms of knowledge which should still be left opaque given this argument. Here, I refer to information regarding the methods of data collation and aggregation used in preparation for the analysis that follows. If adversaries will know how the data is aggregated, they will possibly learn

[249] Stuntz strongly advocates this position, noting that in the context of counterterrorism, where the criminals are well organized and intelligent, "transparency throws out the baby with the bath water." Stuntz, *supra* note 148; *see also* Samaha, *supra* note 150, at 920. For a critical—and even cynical—view of this pro-opacity notion, see COHEN, *supra* note 12, at 207–13.

[250] *See* Erin Murphy, *The Politics of Privacy in the Criminal Justice System: Information Disclosure, the Fourth Amendment, and Statutory Law Enforcement Exemptions*, 111 MICH. L. REV. 485, 503 (2013); *see also* Fenster, *supra* note 152, at 906.

[251] Solove, *supra* note 12, at 361.

[252] COHEN, *supra* note 12, at 256 (making this point while referring to the "Carnival Booth" study).

to leave digital footprints in ways that would be difficult to bring together under one digital roof.

The opacity argument here discussed generates complicated questions while examining its implications for disclosures at the "analysis" stage (designated as B. in Table 1, *supra*), which includes both technological and policy steps. In theory, disclosure as to the forms of analysis carried out might allow for "gaming" the patterns that will emerge in the data-mining process. A related concern is that exposing the software source code will deter firms from developing programs for the government. Firms will fear that such disclosures will reveal trade secrets to their competitors. While both arguments carry a powerful intuitive appeal, I believe they call for additional study (in the fields of both computer science and policy) as to whether indeed the negative outcomes noted will follow. Such studies will seek to establish whether understanding how the automated analysis is carried out allows for its circumvention.

Arguably, this pro-opacity argument could also extend to the "usage" stage (designated as C. in Table 1, *supra*) of the predictive process, starting with the *ex post* studies transparency theories might call for in this context. Feedback from the process, including its rates of success and other attributes might again assist interested parties in sidestepping scrutiny. This pro-opacity argument is relatively unconvincing. Given the importance of sharing this information with the public, such information should be disclosed, unless the government proves that indeed the disclosure of success rates on their own can prove extremely detrimental.

Yet perhaps the most salient context for this pro-opacity argument is elsewhere in the "usage" stage—at the point at which the government uses a mix of criteria, factors, behaviors, and attributes as proxies to identify wrongdoings.[253] Here, the opacity argument is perhaps most intuitive—if government discloses the lists of proxies used, adversaries will simply avoid these proxies. They will, however, still engage in unlawful conduct. Therefore, providing information regarding these steps of the process should be prohibited.[254]

[253] This issue was even addressed by the courts. For instance, in *Gordon v. Federal Bureau of Investigation*, the court blocks a FOIA request for the criteria of the No Fly List. 388 F. Supp. 2d 1028, 1038, 1046 (N.D. Cal 2005). The court noted that it would block such information even if it is publicly known. *Id.* at 1046.

[254] Alongside this argument runs a common counter-response. A great deal of information about governmental strategies could be learned through diligent intelligence gathering and even using proactive trial runs to test the system (a similar argument was mentioned in the text regarding the IRS practices and accountants). Therefore opacity only harms law-abiding citizens, who rely on transparency to achieve the objectives articulated above. Criminals are well aware of the modeling practices, and take various evasive actions to escape them. HARCOURT, *supra* note 43, at 227–31. The "alternative information flows" argument has some merit, but it cannot counter the opacity argument's powerful intuition. After all, not all criminals are sophisticated. Some terrorists act alone. Others are not as smart or are mere copycats. Even sophisticated criminals (luckily) fail at times.

Given the centrality of this pro-opacity argument to the overall discussion regarding transparency in predictive modeling, it calls for some additional discussion throughout the following paragraphs. To do so, we must first distinguish among three forms of proxies for wrongful behavior the government might select at this stage for inclusion in the profile (for the sake of this discussion, let us assume that the process is interpretable and therefore distinguishing among these proxies is possible).[255] At the end of the day, all three forms of proxies should arguably not be disclosed. The strength of the arguments varies, however, and therefore shedding light on this issue is nonetheless important.[256]

One form of proxy refers in itself to *unlawful* conduct. In some cases, the government found these illegal actions to be correlated with the other, more serious crimes it strives to deter (for instance, boarding flights with knives, as a proxy for terrorist activity). [257] For that reason, it flags these illegal actions, seeks them with rigor, and treats those engaging in them with additional suspicion. On its face, arguing for opacity regarding the disclosure of these proxies (the specific forms of illegal behavior) might face an analytical challenge. Disclosure of these factors can lead to social benefits if the impact is avoiding the indicated proxies—those who are planning to carry out the most serious crimes government is striving to predict will sidestep a criminal/antisocial (albeit possibly minor) act so to avoid detection and severe consequences—and in that way disclosure will lead to social benefits. This pro-disclosure argument is not persuasive. The crimes and actions governmental prediction initiatives try to predict are serious and have dire social consequences. Compromising the prediction schemes' effectiveness in order to encourage a few potential criminals to sidestep misdemeanors makes no sense. The minimal benefits of such exposure would be overwhelmed by the potential harms (and lost advantages) that might follow.

Thus, in most cases, maintaining opacity will achieve the governmental objective of lowering antisocial conduct, even after accounting for alternative information flows.

[255] Maintaining the ability to engage in such distinctions is in itself an argument to mandate interpretability.

[256] An interesting argument against opacity, at this juncture, may be that the proxies used to predict crime should be published to educate potential offenders that they might be strolling along a dangerous path they themselves might not be aware of, which others have traveled, with very negative outcomes. It might signal to tax evaders, for instance, to stop claiming a home office when they do not have one. These potential evaders will learn that others have done the same and suffered dire consequences, as this led them to harsher forms of tax fraud. Thus, it might provide individuals with a chance to change their ways.

This interesting notion calls for additional considerations beyond the confines of this Article. Among others, it requires establishing whether it is the role of the state to issue such educational warnings to its citizens, whether these warnings might have unintended negative consequences, and whether they are worth generating an enhanced risk of higher crime and less effective law enforcement, which might follow from such disclosure. I thank Kathy Strandburg for this interesting insight.

[257] SCHAUER, *supra* note 40, at 249.

A second form of proxies pertains to neutral forms of conduct. Nonetheless, these are actions which were found to correlate with the behavior government wants to preempt, reduce, avoid, and regulate.[258] Here, informing the public of these triggers will not have an educational value on its own. Disclosure, however, will allow potential criminals to refrain from generating the triggering event, while continuing to engage in their problematic behavior.[259] For instance, in the tax context, individuals will refrain from claiming a specific true charity deduction, but continue to understate their income. In addition, disclosure might generate unintended negative consequences. Some of the proxies used refer to behavior, which might be socially beneficial (that only launch suspicion when matched with a long set of other factors).[260] Law-abiding citizens will stop engaging in these activities out of fear that they will be subjected to additional scrutiny.[261] For this reason as well, opacity regarding the nature of these proxies is preferable.

Finally, the proxy applied by government might refer to immutable personal factors (or those that are very difficult to change—such as profession or geographical area of residence).[262] In the event of disclosure of the factors used, law-abiding individuals can do little to escape their destiny of being subjected to additional scrutiny (naturally generating frustration and possibly negative stereotypes, as I explain below). But, those who intend to deceive can use forgery and disguise to make the government believe they belong to a group that they do not belong to. In view of the negative implications of stereotyping, this context also seems to call for limited transparency, while other forms of internal accountability must assure that the proxies used are not unfairly discriminatory by nature.

[258] Schauer here refers to the famous example of a broken window and low compliance with safety and sanitary requirements at factories and places of business (and to a theory of causation which can explain why the same personal traits of management might be leading to both). *Id.*

[259] Of course reality generates examples that further complicate the taxonomy here mentioned. For instance, signaling to criminals and terrorists that all cash-based transactions generate suspicion might not deter them but only lead them to use other forms of payment such as debit cards or online banking. These actions, however, might be easier to track, analyze, and later trigger preemption steps. Mapping out all the possible options unfolding at this juncture requires a separate article.

[260] A specific discussion arising in this context pertains to the chilling effects generated by using proxies of behaviors which relate to expression and association (such as attending meetings and reading books). For this discussion, let us assume that these factors are not being used. For more on the specific difficult issues the use of such proxies may raise, see Katherine J. Strandburg, *Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance*, 49 B.C. L. REV. 741 (2008); *see also* Solove, *supra* note 12.

[261] Various examples might come to mind. Some possible situations are avoiding the benefits of traveling without luggage, shopping at discount stores (fearing this might adversely impact their credit rating), having a home office, or even giving large donations to charity.

[262] This discussion raises the important and difficult question as to whether it is fair for governments to distinguish among individuals on the basis of immutable or changeable factors. I will set aside this important discussion for a later time.

In sum, it is apparent that disclosing the inner workings of the predictive process may impede the success of the entire predictive project. This analysis shows, however, that this argument does not apply equally to all stages of the prediction process. Furthermore, it demonstrates the risk associated with allowing for unchecked predictive policies. These might lead to unacceptable outcomes. Therefore, other forms of accountability must be considered when opacity is required. In some contexts, sharing the governmental initiatives with chosen institutions is needed to closely examine whether the governmental actions are prudent.

2. Transparency and Encouraging Crime

Arguably, transparency might also undermine governmental objectives in a slightly different manner. By disclosing the nature of the predictive model, the government is signaling on which segment of the population law enforcement is focused. Therefore, it is also indicating that lower levels of enforcement will be in place for other, much broader, population segments. Such transparency might encourage criminal activity by members of these other population segments, even by individuals who previously refrained from illegal activities.[263] This analytical framework was introduced by Bernard Harcourt as a general argument against the use of predictions (be they automated or manual).[264] Instead, Harcourt advocates a shift to random checks.[265] Harcourt demonstrates this argument while referring to the IRS auditing process, asserting that if the IRS focuses auditing on individuals who meet specific criteria and such information becomes public, individuals who are not part of this group will alter their conduct and cheat on their taxes in greater numbers,[266] as they understand they can do so without being detected. Therefore, (transparent) predictive modeling leads to more, not less, crime. Indeed, knowledge regarding the predictive scheme might increase the crime rate by those outside the government's focus by merely a fraction. Even in such a case, however, the overall outcome of the prediction initiative will be a negative one given the size of the non-scrutinized segment of society.[267]

What Harcourt mostly neglects to address, yet probably takes for granted, are some assumptions regarding information flows concerning

[263] There is a subtle distinction between this pro-opacity argument and the former discussion. The former discussion argues that disclosure will render the prediction model ineffective in stopping the crimes and criminals they are searching for because the latter could use the information provided to sidestep the system. The argument set out here states that transparency will in fact motivate *other individuals* to engage in crime, because they will understand that their actions can now go undetected.

[264] HARCOURT, *supra* note 43, at 2–3.

[265] *Id.* at 5.

[266] *Id.* at 23.

[267] This results from the fact that those within the government's focus are far fewer than those that are outside of it.

these governmental actions. For the problematic dynamic here described to transpire, the public must precisely understand (or at least confidently believe it understands) the way the prediction model is working. Thus, if we adopt Harcourt's assumptions, the problematic outcome he draws out could be countered by effectively *limiting* transparency regarding the predictive process.[268] If the public is unaware of the actual governmental practices, the majority group will not know it is outside of the government's focus. Therefore, these individuals will not alter their behavior, adopt a criminal tendency, and undermine the predictive model. Harcourt's argument against prediction can easily be modeled (probably much to Harcourt's dismay) into a pro-opacity argument. Note, however, that this discussion only refers to transparency in the "usage" stage (designated as C. in Table 1, *supra*) of the prediction process.

This pro-opacity argument, however, which is premised upon Harcourt's model, should be rejected. First, one can challenge the very core of Harcourt's model and its assumptions regarding the high level of elasticity in the individual's willingness to break the law. In other words, law-abiding citizens are not necessarily inclined to break the law even if they know they will not get caught. As an example, Harcourt addresses tax fraud and avoidance—crimes which indeed might carry limited social backlash.[269] Yet in other instances, convincingly making this argument is more difficult (especially in the context of serious crimes and national security, where many automated prediction models are applied).

Second, Harcourt's theory is premised upon the existence of a static prediction model.[270] Automated predictions powered by data mining are a dynamic process, constantly changing in view of new information. Educating the public about the dynamic nature of this process (by providing additional disclosure regarding its inner workings) should not encourage those outside the current scope of law enforcement's focus to engage in criminal activity but rather deter the *entire* population. With such a dynamic process, new trends of criminal behavior would be picked up, and guilty/offending individuals brought to justice. Therefore, the understanding that data mining is applied and how it works should convince individuals contemplating breaking the law to think of something else, rather than bring additional citizens into the realm of crime.

Even though this pro-opacity argument should be rejected, Harcourt's work provides several important lessons for a discussion

[268] Harcourt does not address this matter. HARCOURT, *supra* note 42, at 173. This could either be because he takes for granted the existence of transparency as a requirement which cannot be neglected, or because he assumes that the information will flow to the public anyway. Both assumptions should not be taken for granted.

[269] *See* Blank, *supra* note 42, at 282.

[270] HARCOURT, *supra* note 43, at 223–27.

regarding the role and extent of transparency for automated predictions. It calls for reemphasizing the need for ongoing random checks which must proceed alongside those generated by prediction models. These will provide a baseline for the model's effectiveness, and of course, generate new information for future analysis which will be helpful in identifying individuals shifting toward crime from outside the government's focus. Moreover, it calls attention to the *false negative* factor—those breaking the law while escaping the focus of the predictive model. Harcourt importantly points out that false negatives are not a stable variable.[271] The extent of false negatives might actually grow in view of the nature of the predictive models set in place. Therefore, any transparency scheme must inform the public regularly regarding on the state of false negatives (as part of the *ex post* disclosure requirements detailed throughout this Article), so that a public debate could follow as to the suitability of these predictive measures.

B. Transparency and Stigma

A very different perspective regarding the detriments of transparency in prediction models is derived from problems related to stigma and stereotypes.[272] If the correlations and facts the prediction models are premised upon are to be published (namely, information pertaining to the "usage" stage (designated as C. in Table 1, *supra*)), a very problematic dynamic might follow. The public, who in its majority is untrained in understanding the intricacies of statistical inferences, will rely upon the information published to reach unfair and wrong social conclusions. After reviewing the lists of proxies used in the prediction process, the majority of the public will probably assume that personal traits linked to elevated risk of various anti-social and criminal behaviors are generally indicative of individuals with negative personal traits. While these assumptions are false, they might lead to problematic social outcomes—a process unfolding due to transparency requirements.

The public might misunderstand the meaning of the correlations drawn out and used in the predictive process. First, it will probably misunderstand the power of context.[273] For instance, the correlations published might cause the mistreatment of individuals by their peers in various contexts based on the personal traits they might share with the suspicious pattern.[274] In many of these contexts where mistreatment will

[271] *Id*. at 23 (also referring to the "elasticity" of various segments of the population to reduced law enforcement).

[272] For a general background on stereotypes and the problems they generate, see SCHAUER, *supra* note 40, at 3, 175.

[273] This tendency is commonly referred to as the "Fundamental Attribution Error," and is also recently referred to as the "Correspondence Bias." *See* Daniel T. Gilbert & Patrick S. Malone, *The Correspondence Bias*, 117 PSYCHOL. BULL. 21, 24 (1995).

[274] *Id.* at 21–22.

occur, the correlations uncovered in the predictive process should not be considered relevant. For instance, if data analysis indicates that in specific settings certain law professors writing about privacy and transparency have a higher chance of falsely reporting on their taxes, this does not mean these professors are more likely to engage in non-normative behavior in other contexts—such as act violently or embezzle. Yet this distinction is a delicate point which the public might not comprehend.[275]

Generalizations based on the predictive model can go even further. The public in general might wrongfully interpret correlations as traits about individuals.[276] Rather than understanding a correlation as merely indicating that "people from city *X* have a greater chance of understating their income in IRS filings," they will wrongfully deduce that "people from city *X* are unreliable and untruthful." Therefore, it is fair to assume that if specific factors reappear in correlations applied and published in a variety of governmental contexts, stigmatization of specific groups will follow. An even worse outcome would unfold when the factors used in such correlations are immutable. In such instances, individuals are unable to exit the suspected profile and remain locked into a negative stereotype (thus generating increased frustration).[277]

This discussion of stigma and stereotypes presents familiar problems in a novel, yet nonetheless troubling, setting. Usually, such discussions have focused on specific "protected" groups which have suffered from mistreatment and bigotry for many years.[278] Indeed, many fear that the novel forms of automated prediction analyses will move to discriminate against these groups again.[279] Without reflecting on this point, let us assume (for the sake of this argument) that such outcomes could be avoided. Arguably, it is possible to ensure that none of the patterns used in the predictive process are proxies for the classes of individuals suffering discrimination in the past. Yet even if this was possible, serious

[275] For a recent discussion of the importance of understanding context to predict human behavior (and how predictive behavior changes from one context to another), see YOCHAI BENKLER, THE PENGUIN AND THE LEVIATHAN: THE TRIUMPH OF COOPERATION OVER SELF-INTEREST 66–73 (2011).

[276] Schauer, *supra* note 18, at 2–3.

[277] At this point, the careful reader might state that whether stigma is a problem in this context or not is an interesting question, yet one that surely should not call for opacity in this process. Information regarding the discriminating factors will be sure to leak through alternative information flows anyway. Therefore, the extent of official transparency will not have a substantial impact on the nature of stigma-related concerns. I do not find this statement convincing. Allowing the public to rely on rumors to formulate opinions about individuals is quite different from having an official governmental document to refer to, which maps out governmental approaches to individuals on the basis of personal traits and previous actions. Transparency at this juncture will provide an official stamp of approval for to wrongful personal convictions that would be premised upon an incorrect reading of (nonetheless) real numbers. For that reason, stereotyping is a serious concern to address when considering transparency for automated predictions.

[278] SCHAUER, *supra* note 40, at 175.

[279] *See* SOLOVE, *supra* note 164, at 191.

and novel stigma-related concerns arise. The predictive modeling context inflicts newly formulated groups with stigma-related concerns in a way society has yet to fully understand. Predictive processes might create new forms of social sub-groups which will be subjected to ridicule and mistrust. These population segments, however, might not be divided along well-recognized lines—such as race, gender or nationality. They might also be dispersed socially and geographically. Therefore, they might be unable to mobilize and counter such attacks. All these novel dynamics which society has yet to fully grasp are enhanced by some forms of transparency in the predictive analysis process. For that reason, transparency requirements in this context (the "usage" stage of the predictive process) should be treated with caution.

This novel "stereotyping" concern does not exclusively pertain to the implications of disclosing specific *factors* produced by the prediction analysis and *used* by government. An additional interesting and alarming implication worth discussion is its impact on a possible "causation requirement." As explained above, given the fact that the prediction process is automated by nature, it is possible that it does not inherently include an examination as to the *reasons* specific factors were selected. Thus, a "causation requirement" will mandate that analysts must identify a theory of causation and disclose it every time a correlation is used. Arguably, the stigma-related concern calls for limiting the search for causation in the models applied, or at least mandating opacity for such findings. Formulating a causation theory can potentially do more harm than good. While sound causation might dispel some stereotypes, it might strengthen many others. Causation renders stereotyping more scientific. Therefore, until additional studies establish the social impact of causation disclosure, causation theories should be left for internal review only.

On the other hand, this unique stereotyping concern does not exclusively promote opacity. Quite to the contrary, it promotes transparency at other junctures of the prediction process—especially at the "analysis" stage (designated as B. in Table 1, *supra*). In view of this concern, at the "analysis stage" the government should move to provide additional information as to what levels of correlations and errors it finds acceptable. To battle stigma, it is crucial that the public understands that correlations and other patterns used do not result from direct matches, but from statistical compromises. With such understanding in mind, at least some individuals will refrain from generating, relying upon, or acknowledging stereotypes on the basis of the factors used in governmental prediction (if and when information about these practices becomes public). In other words, a better informed public understanding of the prediction process might limit the logical pitfalls which generate stereotypes, and eventually lead to unfair bias and stigma. Furthermore, stigma-related concerns promote disclosures of *ex post* studies of the data usage. These studies will examine the impact of prediction modeling on minorities and other protected groups. The prospect of these studies' publication will motivate decision-makers to take precautionary steps to limit practices which unfairly discriminate against these protected groups.

VI. ACHIEVING TRANSPARENT PREDICTIONS

Our extensive journey through the realm of transparency in the context of automated prediction is nearing its end. Let us now bring together all the elements discussed and produce a comprehensive set of policy recommendations. Note that a crucial segment of the analysis is still missing and must be filled in when applied in the future to specific contexts and tasks. At that point, the foundations introduced in this Article must be supplemented by an analysis which accounts for the various rights (such as democracy or the facilitation of free speech) which might be relevant at the relevant juncture.

Even so, the recommendations mentioned below can prove beneficial in several ways. In some contexts, they can assist in the creation of comprehensive legal policy architecture "from scratch." Yet as demonstrated above, there is no paucity of laws addressing transparency. To the contrary, the automated prediction context uniquely introduces several contexts with an overlap of an abundance of legal regimes. In such cases, the recommendations set forth could prove helpful in formulating an overall policy response, which could be derived from the proper interpretation of the existing laws and rules (with crucial supplements added where needed).

The conclusions and recommendations set forth below also allow for summing up the theoretical background this Article sets forth. Indeed, this Article set forth an abundance of theories. On their face, these theories of transparency and opacity are, in many cases, conflicted. Yet as the discussion below reveals, they often merely overlap, and allow simple conclusions and recommendations to emerge.

I begin with the recommendations for the "collection" stage (designated as A. in Table 1, *supra*). The lists of datasets applied at this stage should be disclosed. The information within them should not (with the exception of disclosure to the relevant data subjects). These two points are quite simply derived from the analysis above. The technical measures for collating these datasets are best kept outside the public eye, while auditing of this process would be carried out by selected institutions.

The latter part of this conclusion is less convincing, yet follows from accounting for and balancing out the various theoretical discussions noted above, as follows. The pro-transparency theories do not provide strong arguments for broad disclosure of these technical factors; disclosing these technical elements is not likely to generate sufficient public interest to shame lower-level officials into changing their practices or generate a sufficient market-like dynamic. The crowdsourcing argument carries some merit at this juncture. Indeed, the technical decisions made could benefit from outside assistance. The autonomy-based arguments do not provide substantial insights for mandating transparency at this specific juncture. Those premised upon the rights of data subjects (and privacy) might justify additional disclosure of these technical factors. Yet as mentioned, these theories are analytically weak. On the other hand, it is quite a long shot to connect such disclosure requirements at the early stage of the prediction process to the autonomy rights of those affected by the data-mining analysis—concerns arising on the opposite side of the information flow.

These theoretical insights which, to some limited extent, advocate transparency for these specific technical measures must be balanced against the pro-opacity argument, stating that disclosure will undermine governmental initiatives. Those striving to game the law enforcement process will greatly benefit from insights into the aggregation and collation process. They will use such information to understand how they might be able to escape having their information aggregated into one dataset. In view of these concerns, I find that the promise of crowdsourcing and the autonomy requirements should probably be responded to by providing an internal review of the process, while partially relying on selected experts. Therefore, a balancing of the different (and at times, competing) theoretical concerns leads to the specific compromise noted above.

Formulating transparency recommendations for the "analysis" stage (designated as B. in Table 1, *supra*) presents more of an analytic challenge. Currently, the public is left almost entirely in the dark at this stage. This must change. Additional layers of disclosure should be applied to both technological elements and human[280] and policy decisions.

This brings us back to the theoretical justifications. Let us begin with examining disclosures related to *technology* (the software tools enabling the data mining analysis) in view of pro-transparency theories. Here, disclosure generates some modest theoretical benefits. Again, the first "fairness-enhancing" theory is relatively weak. Transparency will only serve as a minimal "check" on governmental actions. The public will have little interest in these technical details. Thus, there is only limited opportunity for effective shaming or affecting political forces. The crowdsourcing argument has greater force. Technical measures could be enhanced by external review. Autonomy-based arguments are

[280] The analysis does not address the transparency requirements for the protocols of decisions and actions made by the relevant bureaucrats at this juncture. As this issue addresses a broad array of actions and decisions, I can merely recommend that the factors addressed throughout this Article be applied so as to engage in balancing the appropriate transparency solution for these practices.

quite a stretch. As detailed above, it is an analytical stretch to link the rights of both "data subjects" and "affected individuals" to the computer analysis process here discussed.

These benefits must be balanced against the unpredictable detriments of exposing the computer code to the public. Such disclosure might allow for manipulating the process and discourage firms from producing software for the government. Given these concerns, a context-specific balance will be required of regulators at specific junctures. A possible compromise calls for releasing the software to a selected group of experts throughout the industry. These experts will be barred from sharing such code commercially. They, however, would be able to inform the public if hidden agendas are imbedded within the software.

Moving to the realm of *policy* decisions, it is important to promote greater transparency with regard to information concerning the support and confidence levels[281] acceptable in the prediction process. Pro-transparency theories clearly point to this conclusion. The internal balances between accuracy and security, which led to the selection of support and confidence levels, will surely generate public interest that will prove to be an effective check on government. These decisions also impact the personal autonomy of those affected by the analysis; with such data in hand they can have a better understanding of the connection between the results of the governmental analysis and their own actions which might have rendered these individuals suspicious.

On the other hand, only limited arguments for opacity hold at this juncture. Disclosing the nature of the support and confidence factors would not undermine the government's objective, as it does not expose the nature of the patterns applied by government. Moreover, stigma and stereotyping concerns should enhance disclosure at this juncture. When the public understands that the correlations and other patterns used are not a result of direct matches, but of statistical compromises, it is possible that it will refrain from generating stereotypes (or at least their negative effects would be limited).

The "usage" stage (designated as C. in Table 1, *supra*) has generated the greatest interest in terms of transparency. Here, we witness various rationales pointing in different directions. Thus, it is important to parcel this stage into its various elements. First, we examine *disclosure of the actual patterns used*. The arguments for transparency are strong; these are matters within the public interest, and both shaming and political forces will be in place. Autonomy interests are also extremely relevant (yet crowdsourcing arguments are relatively weak). Yet, the arguments regarding opacity are of greatest strength here as well, as revealing the actual factors used will allow individuals to

[281] For an explanation of these terms, see *supra* notes 129–31 and related text.

circumvent the governmental objective. In addition, revealing the actual patterns used can promote stereotyping. It might also lead to political pressure to refrain from these practices even when they provide long term benefits.

In addition and practically speaking, when taking into account the existing legal rules and governmental sentiment, calling for transparency in this element has no chance—and probably with good reason. Internal reviews by selected institutions must suffice.

The next point to assess is that of *interpretability*—whether we must require that all relevant processes will be understandable to humans (even if the process is only disclosed internally). I believe such a requirement is crucial, even at the cost of lowering overall efficiency. Engaging in "interpretation," even merely internally, has important implications. For instance, the interpretation process might reveal that in the course of applying automated prediction, ridiculous factors are being applied. Even if the government might choose to continue using this process, it runs the risk that someone might leak the government's specific knowledge (which now exists within the system) of these actions. Thus, the government will think twice before using such correlations. The "interpretation" process is therefore important for generating information, even if it stays within the system.

In addition, carrying out an internal interpretation process will enhance the autonomy of those affected by it. Individuals might not be privy to the "logic" behind the decision that affected them, but will at least know someone is looking into the matter and has additional tools to do so. Interpretability also serves an important role in distinguishing among various proxies. As mentioned above, different groups of proxies lead to different forms of unintended consequences. Government must internally examine the different proxies chosen, while noting the problematic implications of using every one of them (focusing on possible chilling effects and discriminatory outcomes). Again, without interpretation, such distinctions cannot be drawn.

On the other hand, pro-opacity arguments against "interpretability" are relatively weak. Given the fact that the process will be internal, there is a limited security risk in engaging in this process, as well as a limited stigma-related problem. Efficiency of the overall process might suffer given the need to refrain from non-interpretable processes, but this is probably a price we must pay so as to enhance the important objectives mentioned.

In addition, we must address the notion of both mandating and disclosing theories of causation prior to using various factors and proxies in a prediction scheme. Such a "causation requirement" can be derived from pro-transparency theories. Causation will promote effectiveness and act as an important check on governmental actions (given the fact that the nature of these theories will surely generate public interest).

Developing such models, even internally, will enhance autonomy, as they will provide an additional element to assure the process is not arbitrary. Causation studies might also enhance the "crowdsourcing" dynamic (even when only shared with selected institutions). Experts will examine the strength of the causation theory, or try to come up with an alternative one and in that way improve and critique the prediction process.

Causation requirements, however, have their downsides. Disclosure of such theories can open the door to serious privacy and stigma concerns. This might even follow from developing theories and examining them internally, in view of potential leaking of such information to the general public. Given these concerns, great caution must be exercised prior to setting a rule regarding mandatory disclosure of causation requirement. Future analysis must establish the need and importance of internal causation studies.

To summarize our discussion thus far, I acknowledge that readers might find the most recent recommendations disappointing; the theoretical discussion above went to many lengths to examine the importance of transparency at this juncture, while distinguishing between various forms of proxies, and pondering whether we must opt for partial or full disclosure. At the end of the day, however, all these distinctions are assumedly set aside, and an overall recommendation for opacity is advocated. Yet the full theoretical analysis carried out above was not in vain. By understanding the precise reasons for both disclosing and withholding information regarding the actual criteria used, the theoretical framework presented will allow for accurate balances in specific cases which might arise.

For instance, when sidestepping the prediction method will not cause vast social harms, and the factors selected for prediction are not likely to generate serious stereotype concerns (given their innocuous nature), the rules presented above might be changed. For example, if a prediction system is employed to sort out low-level fraud in the allocation of benefits, and the factors used are merely general, mutable, household statistics, the importance of empowering the individuals and providing them with a sense as to why they were treated differently will lead to advocating for overall transparency (including with regard to the nature of the factors used) instead of opacity.

I now conclude with an additional aspect of transparency relevant to the "usage" stage—the drafting and distribution of various *ex post* studies of the prediction process. Several transparency-related arguments address the importance of these measures. The interests in disclosure according to the "efficient policy" theory are especially strong. The information such studies convey will most likely gain public and political traction. Autonomy would also be enhanced if individuals know the process which impacted them is successful overall, and thus worth their personal sacrifice. This information is also crucial to battle concerns that the "classic" forms of stereotyping are crawling into this novel process. On the other hand, information disclosed will not prove helpful in manipulating the system and therefore pro-opacity arguments will mostly be rejected.

On rare occasions, predictive modeling is merely intended to generate deterrence by leading the public to wrongfully believe that the futile tools used are effective. Predictive modeling might be applied as a symbolic tool,[282] meant to assure the public and provide a (false?) sense of security. In the context of national security, such strategy is referred to as "Security Theater."[283] Obviously, *ex post* studies and disclosures will expose these entire practices as a charade. Therefore, publishing the studies mentioned will most probably lead to the termination of the program. Whether the government should engage in these forms of deception to promote important national objectives is a thorny question.[284] The analysis thus far assumed that automated prediction processes are not a form of "Security Theater," and used as an efficient measure for achieving their designated objective (such as finding tax evaders or criminals). Therefore, publishing *ex post* studies should not pose a problem. In very specific and crucial instances, however, applying a "Security Theater" could be acceptable. In these instances, special opacity rules should be created to facilitate such "theaters." These will probably include limiting the publication of the *ex post* studies mentioned. Yet this must be a rare exception and should be addressed separately.

VII. CONCLUSION: THE LIMITS OF TRANSPARENCY

Transparency is hailed as an important policy tool which could enhance autonomy and forward democracy. Its role in the age of information technology has yet to be firmly established. This Article takes initial steps in setting forth a comprehensive mapping for meeting the transparency challenge in a specific context—that of the predictive analysis of personal information.

While acknowledging the important strengths of transparency, it is crucial to recognize there is much harm that governmental prediction

[282] Solove, *supra* note 12, at 352 (stating that this is not the case, but that data mining should be applied only with evidence of effectiveness).

[283] BRUCE SCHNEIER, BEYOND FEAR: THINKING SENSIBLY ABOUT SECURITY IN AN UNCERTAIN WORLD 38–40 (2003); *see* Schwartz, *supra* note 164, at 310–11. For a similar argument, see JEFFREY ROSEN, THE NAKED CROWD: RECLAIMING SECURITY AND FREEDOM IN ANXIOUS AGE 8 (2005).

[284] *See* Adam M. Samaha, *Regulation for the Sake of Appearance* 125 HARV. L. REV. 1563, 1585–90 (2012).

models could generate, and transparency alone cannot cure.[285] For instance, one must question whether allowing the government to obtain a powerful tool, which can generate such insights, is wise. Additional concerns set forth are that the process is ineffective, error-ridden, generates chilling effects, leads to unfair discrimination, and is prone to enable or even encourage function creep. Transparency provides a partial response to all these problems. Additional work must establish how effective a cure it is and what other steps must be taken. For this reason, the analysis here presented is an essential, yet certainly not a final, step.

This blueprint for analyzing the proper role and balance for transparency, however, can serve many other objectives. Government is faced with a variety of novel tasks. Information technology enables information sharing, in real time, with the public. Yet transparency is famous for its unintended consequences. This Article makes an additional modest step in understanding these consequences in the context of predictive modeling, and provides tools for doing so in other contexts as well.

[285] A similar point is made by COHEN, *supra* note 12, at 239 ("[O]perational transparency may be a necessary condition for human flourishing in the networked information society, but it is not a sufficient condition.").