

BIG DATA AND THE RIGHT TO POLITICAL PARTICIPATION

Michal Saliternik*

-Draft-

University of Pennsylvania Journal of Constitutional Law
(forthcoming 2019)

In recent years, the big data revolution has rapidly expanded from the private to the public sector. Today, government authorities at all levels analyze mass amounts of digital data produced by citizens and use it to inform their policy choices in such diverse areas as healthcare, education, transportation, and urban planning. Proponents of this trend assert that it not only yields better policies, but also facilitates political participation by allowing more people to influence governmental decisions at a low cost and with little effort.

This Article argues, however, that the political participation that big data analysis currently enables is flawed in two main respects. First, such participation is usually passive and unintentional, and does not leave room for public deliberation over contested issues. Second, the apparent neutrality of big data may obscure the systematic exclusion of socioeconomically disempowered groups who do not produce digital data that can affect public policy. To explicate these problems, the Article turns to the work of political philosopher Hannah Arendt, especially to her conception of political action and speech and to her idea of the “right to have rights.” It then demonstrates these problems in recent big data initiatives in the fields of healthcare and urban planning.

Finally, the Article asserts that in view of its participatory deficits, big data-based policymaking in its present form may be incompatible with constitutional norms. It argues that under an uncommon yet plausible interpretation, the First Amendment may be understood to establish the positive right of citizens to participate in governmental policymaking in a manner that allows them to express reasoned opinions and engage in public deliberation. It also argues that the Fourteenth Amendment may be understood to establish the right to equal participation in policymaking of all segments of the population, including socioeconomically disadvantaged groups. The Article explains how exactly these alleged constitutional rights apply to big data analysis and discusses some measures that government authorities can take to meet their corresponding obligations without giving up the efficiency advantages of big data-based policymaking.

*Assistant Professor, Netanya Academic College School of Law. I am grateful to Eyal Benvenisti, Yuval Benziman, Michael Birnhack, Tammy Harel Ben-Shahar, Doreen Lustig, Idit Shafran Gittleman, and Yaniv Roznai for their helpful comments and advice.

TABLE OF CONTENTS

INTRODUCTION.....	102
I. THE PITFALLS OF POLITICAL PARTICIPATION THROUGH BIG DATA.....	107
A. MILD PARTICIPATION	107
B. EXCLUSION OF DISADVANTAGED GROUPS.....	109
II. HANNAH ARENDT'S ACCOUNT OF POLITICAL PARTICIPATION	114
A. ARENDT ON POLITICAL ACTION AND DELIBERATION.....	115
B. ARENDT ON POLITICAL EXCLUSION AND THE "RIGHT TO HAVE RIGHTS" ..	117
III. BIG DATA-BASED POLICYMAKING IN PRACTICE	122
A. HEALTHCARE.....	122
B. URBAN PLANNING	128
IV. IS THERE A CONSTITUTIONAL RIGHT TO MEANINGFUL AND INCLUSIVE PARTICIPATION?	132
A. MILD PARTICIPATION AND THE FIRST AMENDMENT	133
B. EXCLUSION OF DISADVANTAGED GROUPS AND THE FOURTEENTH AMENDMENT	139
CONCLUSION	146

INTRODUCTION

In the past five years or so, the big data revolution¹ has rapidly expanded from the private to the public sector. Following in the footsteps of commercial corporations that collect and analyze digital traces left by customers in order to offer better services, develop new products, and provide targeted advertising,² federal and state authorities have begun to collect and analyze digital traces left by citizens in order to improve public policies.³ These digital traces are extracted from various

¹ The term “big data” is commonly used to describe datasets containing massive amounts of information, which is systematically analyzed to “detect patterns, glean insights, and predict answers to complex questions.” *See*, Kate Crawford, *Think Again: Big Data*, FOREIGN POL’Y (May 10, 2013), http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data. The growing use of big data analysis is often described as a revolution not only because it is enabled by advanced technologies, but also because it represents “a shift in mindset about how data could be used,” which is gradually changing the face of many areas of human endeavor. *See* Viktor Mayer-Schönberger & Kenneth Cukier, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* 5-6 (2013).

² *See, e.g.*, BERNARD MARR, *BIG DATA IN PRACTICE: HOW 45 SUCCESSFUL COMPANIES USED BIG DATA ANALYTICS TO DELIVER EXTRAORDINARY RESULTS* 1, 17, 69, 71(2016) (describing commercial uses of big data); danah boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO., COMM. & SOC’Y 662, 675 (2012) (noting that businesses use big data for diverse purposes such as targeted advertising, product design, and traffic planning).

³ *See, e.g.*, Executive Office of the President of the United States, *Big Data Across the Federal Government* (March 2012) [hereinafter Executive Office of the President, *Big Data Across the Federal Government*] (describing federal big data projects); Executive Office of the President of the

sources, including social media sites, internet searches, smartphone applications, and global positioning systems,⁴ and the insights that they generate inform policies in such diverse areas as healthcare,⁵ urban planning,⁶ education,⁷ energy saving,⁸ disaster management,⁹ social welfare,¹⁰ and crime control.¹¹

The growing reliance of government authorities on big data analysis has many advantages. It offers the authorities better knowledge and understanding of the habits and needs of those affected by their policy choices, and allows them to make decisions that are based on updated, comprehensive information.¹² Some commentators argue that big data thereby not only enhances policy outcomes, but also facilitates a new form of mass participation in public affairs, which strengthens

United States, Big Data is a Big Deal (March 2012) [hereinafter Executive Office of the President, *Big Data is a Big Deal*] (describing the National Big Data Research and Development Initiative); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 404-408 (2014) (discussing the use of big data by government institutions); Ron S. Jarmin & Amy B O'hara, *Big Data and the Transformation of Public Policy Analysis*, 35 J. POL'Y ANALYSIS & MGMT. 715, 715 (2016) ("Recent years have seen a growing number of uses of novel 'big data' sources to monitor, improve, and study the delivery of public services.").

⁴ See, e.g., Gema Bello-Orgaz et al., *Social Big Data: Recent Achievements and New Challenges*, 28 INFO. FUSION 45, 45 (2016) (listing various data sources); boyd & Crawford, *supra* note 2, at 663 (same).

⁵ See *infra* Part III.A.

⁶ See *infra* Part III.B.

⁷ See, e.g., Barbara Means et al., U.S. Dep't of Educ., Use of Education Data at the Local Level: From Accountability to Instructional Improvement (2010), www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf (describing the systematic collection and analysis of information generated by and about students as a means to improve educational policies).

⁸ See, e.g., Kathryn Dykes et al., National Renewable Energy Laboratory, Enabling the SMART Wind Power Plant of the Future Through Science-Based Innovation (Aug. 2017), www.nrel.gov/docs/fy17osti/68123.pdf (describing the Department of Energy's Atmosphere to Electrons (A2E) program, which explores ways to improve wind power plants activity by using innovative technology, including big data analysis).

⁹ See, e.g., Executive Office of the President of the United States, The Federal Big Data Research and Development Strategic Plan 8 (2016) [hereinafter, Executive Office of the President, The Federal Big Data Research and Development Strategic Plan] (noting that the Department of Homeland Security together with the Federal Emergency Management Agency and the US Fire Administration are using big data technologies to deliver new insights on how to improve emergency training and reduce losses).

¹⁰ See, e.g., Erika M. Kitzmiller, *IDS Case Study: Allegheny County's Data Warehouse: Leveraging Data to Enhance Human Service Programs and Policies* (2014), www.aisp.upenn.edu/wp-content/uploads/2015/08/AlleghenyCounty-_CaseStudy.pdf (describing how the Department of Human Services in Allegheny County, Pennsylvania, aggregates and analyzes data from multiple sources to improve its services).

¹¹ See, e.g., Andrew Guthrie Ferguson, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 4 (2017) (describing how police departments throughout the country use big data analytics to predict and investigate criminal activity).

¹² See, e.g., Roger Stough & Dennis McBride, *Big Data and U.S. Public Policy*, 31 REV. POL'Y RES. 339, 339 (2014) ("analysis of 'Big Data' offers potentially to provide public sector policy makers with extensive new information that would inform policy at unprecedentedly detailed levels.").

democracy, empowers citizens, and enhances the legitimacy of government actions.¹³ According to this view, the political participation that big data makes possible improves on other methods of political participation—including notice and comment procedures, closed discussions with relevant stakeholders, public hearings, and open online consultations—in that it engages broader segments of the population in government decision-making at relatively low cost.¹⁴

However, despite these potential advantages, political participation through big data suffers from two main flaws.¹⁵ First, governmental big data analysis usually relies on the reuse and re-adaptation of data that citizens originally produced for other purposes.¹⁶ In these circumstances, the contribution of citizens to the policymaking process is passive and indirect, and they do not have sufficient control over the determination of their needs and preferences and their translation into policy choices. Moreover, even in cases where citizens produce digital data specifically for policymaking purposes and communicate it directly to the government authority through a designated platform, they hardly have an opportunity to express their views about the policy issues at stake, let alone engage in reasoned discussion about them. Their participation in the collective decision-making process thus remains technical and schematic. This type of civic participation in public affairs is referred to in this Article as “mild participation,” as opposed to deep, deliberative forms of political participation.¹⁷

Second, participation in big data-based policymaking depends on the production of relevant digital data. This means that people with lower income and less education—who, according to empirical studies, produce less digital data than

¹³ See, e.g., Thomas M. Philip et al., *A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles*, 18 TECH., KNOWLEDGE, & LEARNING 103 (2013) (suggesting that basic big data literacy facilitates democratic participation); Matthew Tenney and Renee Sieber, *Data-Driven Participation: Algorithms, Cities, Citizens, and Corporate Control*, 1(2) URBAN PLANNING 101 (2016) (describing people’s contribution to large datasets of geographic information as a form of (imperfect) civic participation).

¹⁴ See, e.g., Jonathan Bright & Hellen Margetts, *Big Data and Public Policy: Can It Succeed Where E-Participation Has Failed?* 8 POL’Y & INTERNET 218, 219 (2016) (asserting that the use of big data in policymaking can offer a solution to the problems of low participation and high costs associated with many online consultation initiatives); Eyal Benvenisti, *Upholding Democracy Amid the Challenges of New Technology: What Role for the Law of Global Governance?* 29 EUR. J. INT’L L. 671 (2018) (“The availability of big data and the fast and relatively cheap means to process it are prompting public and private governance bodies to regard the traditional bidirectional communications process as unnecessarily burdensome, if not superfluous.”).

¹⁵ Of course, big data-based policymaking suffers from many limitations, ranging from poor quality of data to data manipulations to possible infringements on privacy and other human rights (see, e.g., *infra* notes 50-53 and accompanying text). This Article, however, focuses on the problems directly associated with the impact of big data on the right to political participation.

¹⁶ See *infra* note 27 and accompanying text.

¹⁷ See *infra* Part I.A.

others—might miss important opportunities to participate in public affairs.¹⁸ Moreover, since big data is commonly believed to be neutral, objective, and apolitical, this systematic exclusion can easily go unnoticed and be left unaddressed.¹⁹ Hence, governmental reliance on big data analysis might create an ‘illusion of inclusion,’ while actually entrenching socioeconomic and political gaps.²⁰

One might think that although the shortcomings of participation in public affairs through big data analysis—namely, its mild nature and the exclusion of disadvantaged socioeconomic groups—are disturbing, they do not present such an urgent problem. After all, many governmental decisions are still being made without resorting to big data, and other methods of political participation such as public comments on suggested regulations and online consultations on various policy questions remain pervasive. However, this reality seems to be changing very rapidly. Taking into account the great interest of U.S. government agencies in big data opportunities,²¹ as well as the global trend towards big data-based policymaking,²² it seems safe to assume that the big data revolution in the public sector is only beginning, and it is here to stay.²³

In view of this prediction, it is necessary to acknowledge and understand the pitfalls of political participation through big data and to examine possible ways to address them. In Part I, the Article begins this exploration by elaborating on the sources and implications of the problems of mild participation in and exclusion from big data-based policymaking. Part II turns to the work of political philosopher Hannah Arendt to offer a deeper theoretical account of these deficiencies. It discusses Arendt’s conceptualization of deliberate action and opinion as the fundamental building blocks of political participation, as well as her claim that every person should have the right to such political participation, which she defines as the “right to have rights.” While Arendt developed these ideas to explain the ramifications of the state-centered political order of the mid-twentieth century, this Article suggests that with some adaptations, they can also be invoked to explore the

¹⁸ See *infra* Part I.B.

¹⁹ See *infra* Part I.B.

²⁰ See *infra* Part I.B.

²¹ This interest has been stated clearly in a series of big data studies and reports published by the federal government. See, e.g., EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA ACROSS THE FEDERAL GOVERNMENT, *supra* note 3; EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA IS A BIG DEAL, *supra* note 3; EXECUTIVE OFFICE OF THE PRESIDENT, THE FEDERAL BIG DATA RESEARCH AND DEVELOPMENT STRATEGIC PLAN, *supra* note 9.

²² See, e.g., Gang-Hoon Kim et al., *Big Data Applications in the Government Sector*, 57(3) COMMUNICATIONS OF THE ACM 78, 83-84 (2014) (describing the increasing use of big data applications by the governments of technologically advanced countries).

²³ See, e.g., *Big Data and the Public Sector: A Survey of IT Decision Makers in Federal and State Public Sector Organizations*, TECHAMERICA FOUNDATION (Feb. 2014), www.splunk.com/pdfs/fact-sheets/sap-public-sector-big-data-report-final-2.pdf (reporting that 82% of information technology decision makers in the federal and state governments expect that the use of big data in the public sector is “the way of the future”).

implications of big data-based policymaking in the globalized, datafied world of the twenty-first century.

Part III demonstrates the problems of mild participation in and exclusion from big data-based policymaking in two preeminent domains of the public sector's big data revolution, namely, healthcare and urban planning. It details several governmental big data initiatives from recent years and shows that while these initiatives promote policymaker consideration of citizens' objective health conditions and urban experiences, they do not offer citizens the opportunity to present and discuss their subjective preferences and views about related policy issues. It also shows that due to various technological, financial, and mental barriers, members of socioeconomically disadvantaged groups are less likely to contribute data to these projects. This means that those who are in greatest need of public health and urban services may have the fewest opportunities to affect them.

Part IV argues that in view of its participatory deficits, big data-based policymaking in its present form may be incompatible with constitutional norms. It notes that even though the right of citizens to participate in public affairs is not explicitly embedded in any provision of the Constitution, it may be understood to underlie the entire Constitution. Drawing on a late-twentieth-century body of scholarship that points to the influence of civic-republican ideals on the Framers of the Constitution, Part IV suggests that the First Amendment can be interpreted as placing upon the government a positive duty to provide adequate opportunities for people to engage in meaningful discussions over public affairs. Hence, to the extent that the government relies in its decisions on the production, collection, and analysis of big data, it must make sure that this process is as deliberative as possible. One possible way to do this is to employ, and push forward the development of, advanced artificial intelligence technologies that allow for the mass scale automatic analysis of complex texts.

Part IV further examines whether the exclusion of socioeconomically disadvantaged groups from governmental big data analysis could be deemed unconstitutional under the Fourteenth Amendment, either through the Equal Protection Clause or through the Privileges or Immunities Clause. It notes that under the prevailing disparate treatment interpretation of the Equal Protection Clause, exclusion from big data analysis is likely to survive judicial scrutiny. For one thing, it is hard to associate such exclusion with a discriminatory intention. For another, socioeconomically disadvantaged people are generally not considered a protected class. However, if the Supreme Court ever accepts that the Equal Protection Clause also prohibits unintentional disparate impact—as many commentators assert it should—big data-based policies that disfavor socioeconomically disadvantaged populations may be deemed unconstitutional.

The Fourteenth Amendment's Privileges or Immunities Clause has gained far less attention and has been invoked far more rarely than the Equal Protection Clause, mainly because nineteenth-century judicial decisions interpreted it so narrowly that it became meaningless. In recent years, however, legal scholars have convincingly argued that this Clause should be interpreted more broadly as

securing, *inter alia*, the right to equal and effective participation in political life. This interpretation seems to entail that all citizens should be able to equally contribute to governmental big data analysis. Part IV concludes by suggesting concrete measures that government authorities can take in order to meet their purported duty under either the Equal Protection Clause or the Privileges or Immunities Clause to promote big data inclusiveness. The Conclusion summarizes the discussion.

I. THE PITFALLS OF POLITICAL PARTICIPATION THROUGH BIG DATA

A. *Mild Participation*

When public authorities engage in big data analysis, they can use two types of digital data. The first type, which may be called “non-policy-oriented big data,” is data that was originally created for other purposes and which happens also to be useful for policymaking purposes. Such data is usually produced without any intention to affect public policy and with no reflection on its possible policy implications. Examples of non-policy-oriented big data include patient medical records that federal agencies collect from hospitals and then analyze to improve public health services,²⁴ as well as driver-to-driver communications that the departments of transportation in several states draw from navigation applications and then use to improve road planning and infrastructure.²⁵

The second type of data that government authorities use, which may be called “policy-oriented big data,” is created with the knowledge and specific intention that it be collected and analyzed for particular policymaking purposes. Examples of this type include reports about hazards, nuisances, and other disturbances that city residents communicate directly to their local authorities through various online applications. These reports are intended to help the authorities solve immediate problems as well as improve urban systems in the longer term.²⁶

Most of the big data that public authorities currently use belong to the first, non-policy-oriented type.²⁷ While the analysis of such data can provide decisionmakers with important insights concerning the habits and needs of those affected by their choices, it does not give the latter sufficient control over the translation of these

²⁴ See *infra* notes 131, 135-136 and accompanying text.

²⁵ See *infra* notes 153-154 and accompanying text.

²⁶ See *infra* notes 157-159 and accompanying text.

²⁷ There seems to exist no comprehensive documentation of the big data sources currently used by the government. However, in view of the fact that designated governmental big data platforms are a relatively new phenomenon, and that most collectable digital data is currently produced under the auspices of commercial companies, it seems safe to assume that most of the data that the government uses is non-policy oriented. It is noteworthy that some of the data analyzed by public authorities has been produced for multiple purposes, both policy- and non-policy-oriented. However, in order to simplify the discussion, data sources are divided here into only two types.

habits and needs into policy choices.²⁸ It allows for the realities and experiences of large segments of the population to be represented in the policymaking process, but it does not give those represented an opportunity to sound their voices and express their perspectives on the policy questions at stake. According to Nick Couldry, such denial of voice and opinion undermines people's humanity and agency, understood as the ability to make rational decisions and be accountable for them.²⁹

Moreover, participation that is based on the secondary use of non-policy-oriented digital data does not allow for public deliberation over contested issues, which arguably represents the ultimate form of political participation. The idea of deliberative democracy suggests that in order to be legitimate, public decision-making must represent something more than the simple aggregation of individual interests or preferences;³⁰ it must be the outcome of authentic deliberation among those affected, which includes a joint search for solutions to collective problems through the exchange of reasoned arguments and proposals.³¹ Such deliberation enhances participants' agency and self-realization,³² and at the same time yields better decisions.³³ It cannot take place, however, when the contribution of people to the collective decision-making process is unintentional and perhaps even unconscious.

These problems are somewhat mitigated when the government uses big data of the second type, which people create with the clear intention of informing policymaking processes. However, even in these cases, participation usually falls

²⁸ As Nick Couldry puts it, big data analysis "aggregat[es] action fragments from any moment in the stream of a person's recorded acts into patterns that bear little relationship to how those people themselves understand the sequence and meaning of their actions." See, Nick Couldry, *A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World*, 67 SOC. REV. 880, 889 (2014).

²⁹ NICK COULDRY, WHY VOICE MATTERS: CULTURE AND POLITICS AFTER NEOLIBERALISM 1, 8 (2010); Nick Couldry & Alison Powell, *Big Data from the Bottom Up*, BIG DATA & SOC'Y 1 (July-Dec. 2014).

³⁰ See, e.g., Cass R. Sunstein, *Beyond the Republican Revival*, 97 YALE L.J. 1539, 1548-49 (1988) [hereinafter Sunstein, *Beyond the Republican Revival*] (noting that the deliberative approach to politics assumes that public discussion can transform and refine individual preferences, which should not be taken as exogenous to politics).

³¹ See, e.g., ARCHON FUNG, EMPOWERED PARTICIPATION: REINVENTING URBAN DEMOCRACY 4 (2004) (describing deliberative decision-making as "a process of structured reasoning in which [people] offer proposals and arguments to one another."); Joshua Cohen, *Deliberation and Democratic Legitimacy*, in THE GOOD POLITY: NORMATIVE ANALYSIS OF THE STATE 17, 21 (Alan P. Hamlin and Phillip Pettit eds., 1989) (asserting that citizens regard their democratic institutions as legitimate insofar as they establish an appropriate framework for public deliberation).

³² See, e.g., Cass R. Sunstein, *Interest Groups in American Public Law*, 38 STAN. L. REV. 29, 36 (1985) [hereinafter Sunstein, *Interest Groups*] (noting that under the antifederalist approach, public deliberation is considered to provide "a kind of 'happiness' that [can] be found nowhere else").

³³ JÜRGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY 302-314 (William Rehg trans., MIT Press 1996) (1992) (contending that democratic legitimacy stems from the combination of formal and informal public deliberation that enhances rational policymaking).

short of being sufficiently deliberative and empowering. Most of the information that people transfer to government authorities for big data analysis purposes is technical, and sometimes it is generated in a semi-automatic way (for example when people operate a cellular application that automatically reports certain activities). This kind of data sharing does not leave room for the expression of complex views or for substantive discussion. It gives people the opportunity to directly affect public policy, yet this influence does not amount to deep engagement in the democratic process.

Hence, whether participation in public affairs through big data is intentional or unintentional, direct or indirect, it seems fair to describe it as a soft, latent, or “mild” form of political participation. The question therefore arises whether big data platforms, whose main feature and advantage is their high volume, could ever be more deliberative and empowering than they currently are, or the increase in quantity essentially comes at the expense of quality. The answer to this question depends to a large degree on the availability of advanced artificial intelligence technologies that allow for rapid and reliable analysis of complex expressions on a large scale.³⁴ Such technological developments, in turn, may be affected by political and legal incentives of the kind discussed in the following Parts of this Article.

B. Exclusion of Disadvantaged Groups

Deliberative models of political participation emphasize not only that people should have appropriate opportunities to express their views and exchange reasoned arguments concerning collective affairs, but also that these opportunities must be equal, so that all those affected can take part in the process.³⁵ This brings us to the second problem with political participation through big data, namely that some people in society produce far less digital data than others and therefore make a smaller contribution, if any, to big data-based policymaking. In particular, digital data is scarcer when it comes to people with lower income and less education, as well as elderly people. Although these characteristics correlate to some degree with race, ethnicity, and gender, empirical studies suggest that the “digital divide” revolves mainly around level of income and education, or socioeconomic status.³⁶

³⁴ See *infra* note 210 and accompanying text.

³⁵ See, e.g., Cohen, *supra* note 31, at 21 (holding that free deliberation among equals is the basis of democratic legitimacy); Jürgen Habermas, *Further Reflections on the Public Sphere*, in HABERMAS AND THE PUBLIC SPHERE 421, 449 (Craig Calhoun ed., 1992) (asserting that deliberation requires “the inclusion of all parties that might be affected”); Seyla Benhabib, *Toward a Deliberative Model of Democratic Legitimacy*, in DEMOCRACY AND DIFFERENCE: CONTESTING THE BOUNDARIES OF THE POLITICAL 67, 70 (Seyla Benhabib ed. 1996) (arguing that legitimate public deliberation must be governed by the principle of equality).

³⁶ For example, a comprehensive survey conducted by the Pew Research Center shows that in 2016, 33% of U.S. adults with less than a high school education did not use the internet, 23% of adults with only high school education did not use the internet, and only 6% of adults with some college education did not use the internet. At the same time, 23% of adults from households earning less

In fact, in the context of big data-based policymaking the problem of the digital divide is twofold. First, members of socioeconomically disadvantaged groups have less access to digital devices, technologies, and networks than members of more privileged groups.³⁷ Hence, many poor, uneducated people do not meet the basic physical conditions necessary for generating digital information. Second, even when they do have digital access, people located low on the socioeconomic scale do not always have the mental resources that enable the creation of digital content that can be relevant for public policymaking.³⁸ Digital activities such as navigating, searching online information, or reporting hazards through designated government applications require a certain sociocultural orientation as well as some technical skills, which less educated people may not possess.³⁹ As others have observed, the digital divide in the Web 2.0 era is based not only on simplistic notions of access, but also on the ability to use technology in meaningful ways.⁴⁰

than \$30,000 a year did not use the internet, whereas less than 3% of adults from households earning more than \$75,000 a year did not use the internet. In addition, 22% of rural adults and 12% of urban adults did not use the internet. Finally, 41% of Americans aged 65 and older did not use the internet, compared with only 1% of 18- to 29-year-olds. The survey also shows that the gaps between men and women as well as between Whites, Blacks and Hispanics are much milder, with 15% of women and 16% of Blacks and Hispanics not using the internet, compared to 12% of men and 13% of Whites. See Monica Anderson & Andrew Perrin, *13% of Americans Don't Use the Internet: Who Are They?* PEW RESEARCH CENTER, (Sept. 7, 2016), www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they (summarizing the findings of the Pew survey). According to the Pew survey, the same variables (educational attainment, household income, community type, and age) also correlate with ownership of mobile devices as well as with the usage of social media (Facebook, Instagram, Pinterest, LinkedIn, and Twitter). See *Mobile Fact Sheet*, PEW RESEARCH CENTER (Feb. 5, 2018), www.pewinternet.org/fact-sheet/mobile and *Social Media Fact Sheet*, PEW RESEARCH CENTER (Feb. 5, 2018), www.pewinternet.org/fact-sheet/social-media.

³⁷ See *id.*

³⁸ See, e.g., Jen Schradie, *The Trend of Class, Race, and Ethnicity in Social Media Inequality*, 15 INFO., COMM. & SOC'Y 555, 567 (2012) (emphasizing the need to acknowledge the class-based constraints that are unique to active production, as opposed to passive consumption, of digital content); Christian Pieter Hoffman et al., *Content Creation on the Internet: A Social Cognitive Perspective on the Participation Divide*, 18 INFO., COMM. & SOC'Y 696 (2015) (arguing that young, highly educated males produce more digital content than others because of their higher self-efficacy and lower concern for privacy).

³⁹ For example, Hargittai and Hinnant show that people with higher levels of education and of a more resource-rich background are likely to use the Web for more "capital-enhancing" activities—such as arranging for travel, seeking health information, researching products, and banking—whereas those with less education and income are more likely to use the internet for entertainment. Eszter Hargittai & Amanda Hinnant, *Digital Inequality: Differences in Young Adults' Use of the Internet*, 35 COMM. RES. 602 (2008). Arguably, the activities that Hargittai and Hinnant define as "capital-enhancing" are more likely to be used for big data analysis than the activities they classify as entertaining, which means that according to their findings, educated and wealthy people have better chances to be represented in such analysis. See also Cary Coglianese, *Citizen Participation in Rulemaking: Past, Present, and Future*, 55 DUKE L.J. 943 (2006) (arguing that the main barriers to citizens' participation in administrative decision-making are cognitive and motivational).

⁴⁰ Michael Crutcher & Matthew Zook, *Placemarks and Waterlines: Racialized Cyberscapes in Post-*

Some scholars contend that the digital divide is diminishing as cellular and internet services become more affordable and younger people who were born into the digital era replace older generations.⁴¹ A few of them invoke the “diffusion of innovations” theory to claim that information and communication technologies are likely to reach near-universal use that crosses all sectors of society, as did the television and (fixed) telephone in their time.⁴² However, the twofold nature of the digital divide suggests that the rumors of its death may be premature. Unlike the television or telephone, contemporary digital technologies are emerging and changing rapidly, thereby maintaining a persistent gap between those who have the skills and resources that facilitate adaptation and those who do not.⁴³ It seems, then, that unless the underlying causes of the digital divide are effectively addressed, it will continue to exist and bear implications for those who are left behind.

One might further argue that even if big data-based policymaking may exclude disadvantaged groups who do not produce relevant digital data, it is still more inclusive of such groups than other policymaking methods, and that it should therefore be celebrated rather than criticized by those concerned about participation in politics. Admittedly, when policymakers make decisions on the basis of their own intuition or judgment, yield to the pressures of strong interest groups, or follow

Katrina Google Earth, 40 GEOFORUM 523, 533 (2009). See also Paul Dimaggio et al., Digital Inequality: From Unequal Access to Differentiated Use, in SOCIAL INEQUALITY 355 (Kathryn M. Neckerman ed. 2004) (emphasizing that social factors affect not only formal access to the internet but also the ability of people to use such access in a meaningful and effective way).

⁴¹ See, e.g., CASS R. SUNSTEIN, REPUBLIC.COM 2.0 17 (2007) (noting that “in both the domestic and the international context, that problem [the digital divide] seems likely to diminish over time, as new technologies, above all the Internet, are made increasingly available to people regardless of their income or wealth.”); Andrew Power, *Governance, Social Media, and the Cybercitizen – Always in Motion is the Future*, 10 SCRIPTED 231, 238 (2013) (asserting that the digital divide based on age is diminishing).

⁴² According to the diffusion of innovations theory, the process of the social adoption of technological innovations often begins with a small, privileged group of “innovators” and “early adopters,” and then moves on to “early majority,” “later majority,” and “laggards.” See EVERETT M. ROGERS, DIFFUSION OF INNOVATIONS 282-285 (5th ed. 2003). Applied to the context of ICTs, this theory leads techno-optimists to believe that, given that access to computing and communication technologies is “growing with a record speed,” the digital divide is a temporary phenomenon. See Ilkka Tuomi, DIGITAL DIVIDES IN THE KNOWLEDGE SOCIETY 1 (background paper prepared for UNESCO World Report on Knowledge Societies, Oct. 20, 2003).

⁴³ See Sabina Lissitsa & Azi Lev-On, *Gaps Close, Gaps Open: A Repeated Cross-Sectional Study of the Scope and Determinants of the Ethnic Digital Divide*, 7 INT. J. ELECTRONIC GOV. 56 (2014) (showing that even when access differences between groups diminish, a ‘second-level’ digital divide of social media usage may persist); Jan A.G.M. van Dijk, *Digital Divide Research, Achievements and Shortcomings*, 34 POETICS 221 (2006) (asserting that while in terms of physical access the digital divide seems to be closing, in terms of digital skills and the use of applications, the divide persists or is widening); Crutcher & Zook, *supra* note 40, at 533 (contending that the digital divide should be seen as “a dynamic and constantly shifting gap in the use of digital resources caused by structural issues that can not be successfully bridged by technology alone.”).

some other decision-making logic that is insulated from public scrutiny,⁴⁴ they are more likely to disregard the interests of disempowered groups than when they rely on big data analysis.⁴⁵ Moreover, even when policymaking involves “traditional” participation mechanisms such as notice and comments, public hearings, or online consultations, it is likely to be less inclusive than big data-based policymaking, which allows many more people to contribute to the decision-making process.

The problem, however, is that big data can easily create an ‘illusion of inclusion’ that makes the exclusion of disempowered groups go unnoticed.⁴⁶ As some scholars have noted, many people believe that big data is inclusive and neutral and that it offers “objective and universal insights into patterns of human behavior.”⁴⁷ Of course, decision-makers who rely on big data may have an interest to cultivate this perception of big data,⁴⁸ as it helps them obscure or at least depoliticize the disregard of vulnerable groups.⁴⁹ Arguably, such misconceptions make

⁴⁴ See, e.g., Coglianese, *supra* note 39, at 945 (noting that “[p]rior to the advent of modern information technology, unelected regulatory officials made significant policy decisions through a process largely insulated from the general public.”); Daniel Esty & Reece Rushing, *The Promise of Data-Driven Policymaking*, ISSUES IN SCIENCE & TECHNOLOGY 67, 69 (Summer 2007) (noting that “[i]n the absence of good data, policymaking frequently relies on intuition, past experience, or expertise...”, and all are susceptible to cognitive and emotional biases).

⁴⁵ See, e.g., Larry M. Bartels, *Economic Inequality and Political Representation*, in THE UNSUSTAINABLE AMERICAN STATE 167 (Lawrence Jacobs & Desmond King eds. 2009) (noting that wealthier and better-educated citizens are more likely to vote, have well-informed preferences, and have direct contact with public officials); Martin Gilens, AFFLUENCE AND INFLUENCE: ECONOMIC INEQUALITY AND POLITICAL POWER IN AMERICA (2012) (detailing the increased responsiveness of American politicians toward the political concerns of affluent citizens).

⁴⁶ The problem of the ‘illusion of inclusion’ was identified in the context of public participation in governmental decision-making through public hearings, written comments, etc. It has been argued that such participation may pay lip service to democratic values while actually having little impact on decision-makers’ choices. See, e.g., Roger Few et al., *Public Participation and Climate Change Adaptation: Avoiding the Illusion of Inclusion*, 7 CLIMATE POL’Y 46, 53 (2007) (observing “a common tendency for pre-existing power relations to persist in participatory fora despite the claims that they promote bottom-up decision-making.”); Judith E. Innes & David E. Booher, *Reframing Public Participation: Strategies for the 21st Century*, 5 PLANNING THEORY & PRACTICE 419 (2004) (noting that the methods of participation in governmental decision-making prescribed by U.S. laws—in particular public hearings, review, and comment procedures—fail to achieve genuine participation).

⁴⁷ See Crawford, *supra* note 1. See also Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2008), www.wired.com/2008/06/pb-theory (“With enough data, the numbers speak for themselves.”).

⁴⁸ See, e.g., Jen Schradie, *Big Data Not Big Enough? How the Digital Divide Leaves People Out*, MEDIASHIFT (July 31, 2013), <http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/> (arguing that in the past, those who focused on elite groups did not pretend that they represented all of society, whereas “[u]sers of Big Data . . . imply that this information does include all ‘citizens.’”).

⁴⁹ See, e.g., Lawrence Joseph with Frank Pasquale, *Interview on the Black Box Society*, BALKINIZATION (Sept. 19, 2004), https://balkin.blogspot.com/2014/09/interview-on-black-box-society_19.html (asserting that big data “is touted as a way to understand and control society without

governmental reliance on big data no less dangerous to democratic values than other, less inclusive, decision-making methods whose exclusionary elements can at least be more easily discerned and criticized.

Another factor that may draw public attention away from the problem of exclusion from big data is the preoccupation of lawyers, scholars, and human rights activists with the risks that big data analysis poses to those who are *included* in digital databases, especially the risks of invasion of privacy,⁵⁰ mass surveillance,⁵¹ cybercrime,⁵² and algorithmic discrimination.⁵³ It seems that the intensive interest of big data critics in the rights of people who are caught in the big data net has led them to overlook the deprivation of those who are left behind.⁵⁴ However, in view

reference to the history (or patterns of thought) that gave rise to the data analyzed.”).

⁵⁰ See, e.g., MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 156 (2013) (asserting that “[i]n the era of big data, the … core strategies long used to ensure privacy . . . have lost much of their effectiveness.”); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11(5) NW. J. TECH. & INTELL. PROP. 239 (2013) (suggesting regulatory measures to address privacy concerns arising from big data); PRESIDENT’S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE (May 2014), www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy--_may_2014.pdf (noting that traditional technologies used to protect privacy can no longer be relied upon).

⁵¹ See, e.g., boyd & Crawford, *supra* note 2, at 664 (contending that big data can be seen “as a troubling manifestation of Big Brother”); MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 150 (noting that everyday activities like the use of credit cards or cellular phones put people under constant surveillance).

⁵² See, e.g., Charles Duhigg, *Bilking the Elderly, With a Corporate Assist*, N.Y. TIMES (May 20, 2007), www.nytimes.com/2007/05/20/business/20tele.html (reporting that companies specializing in compiling big data sold personal financial information to “known lawbreakers”); Bruce Schneier, *Stealing Fingerprints*, VICE: MOTHERBOARD (Sept. 29, 2015), <https://motherboard.vice.com/read/stealing-fingerprints> (discussing the possible implications of fingerprint theft); IDENTITY CARE, NEW ZEALAND: OBSERVATIONS OF IDENTITY COMPROMISE AND MISUSE (2016), www.idcare.org/wp-content/uploads/2016/12/NZ-IDCARE-REPORT-2016.pdf (reviewing misuses of personal identity in New Zealand);

⁵³ See, e.g., Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CAL. L. REV. 671 (2016) (arguing that algorithm-based hiring decisions might discriminate against protected groups); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1191-1205 (2017) (discussing the discriminatory potential of the use of *machine-learning algorithms* by administrative agencies); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017) (presenting a technological toolkit to help ensure that automated decision-making complies with basic standards of fairness and antidiscrimination).

⁵⁴ There are, however, some shorter pieces that discuss the problem of exclusion from Big Data. See, e.g., Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55 (2013) (discussing the threats that Big Data poses to those who remain at its periphery); Schradie, *supra* note 48 (highlighting the “Big Data Gap” and its relation to socioeconomic status). See also CHARLY GORDON, *BIG DATA EXCLUSIONS AND DISPARATE IMPACT: INVESTIGATING THE EXCLUSIONARY DYNAMICS OF THE BIG DATA PHENOMENON* (2014) (unpublished MSc dissertation, London School of Economics and Political Science) (on file with author) (exploring digital exclusion in the private

of the expansion of big data analysis to the public sector, which means that big data outsiders may be denied not only market opportunities but also the ability to influence public affairs, it is high time to take this problem seriously and find ways to address it.

II. HANNAH ARENDT'S ACCOUNT OF POLITICAL PARTICIPATION

In the opening of *The Human Condition*, political theorist Hannah Arendt offers her reflections on scientific and technological innovations and their relation to political life. Writing in 1958, Arendt describes the major scientific achievements of her time—the launch of the space satellite Sputnik I, the splitting of the atom, and the fertilization of a human egg in a test tube—as clear manifestations of “the tremendousness of human power and mastery.”⁵⁵ At the same time, though, Arendt identifies the dangers associated with these achievements. Her main concern is that the ‘truths’ of modern science “may no longer lend themselves to normal expression in speech and thought.”⁵⁶ Since speech and thought are the features that distinguish human beings from other animals and define them as political creatures,⁵⁷ their loss might turn people into no more than human animals,⁵⁸ who live “at the mercy of every gadget which is technically possible.”⁵⁹

Science and technology have developed immensely since Arendt wrote *The Human Condition*, but her observations seem to be as relevant as ever. In the context of the big data revolution, these observations suggest that while governmental deployment of big data technology can greatly enhance administrative efficiency and productivity, it may also alienate people who do not speak big data’s “language” from the domains it governs. This alienation may be reflected in mild public participation in these domains, or even in no participation at all. Section A connects Arendt’s theory to the first risk, and Section B to the latter.

sector).

⁵⁵ HANNAH ARENDT, THE HUMAN CONDITION, 1 (2d ed. 1998) (1958) [hereinafter ARENDT, THE HUMAN CONDITION].

⁵⁶ *Id.* at 3.

⁵⁷ See HANNAH ARENDT, THE ORIGINS OF TOTALITARIANISM 297 (New ed. Shoken Books, 2004 (1951)) [hereinafter ARENDT, THE ORIGINS OF TOTALITARIANISM] (noting that “man, since Aristotle, has been defined as a being commanding the power of speech and thought . . . and . . . has been thought of as the ‘political animal,’ that is one who by definition lives in a community”). See also ARENDT, THE HUMAN CONDITION, *supra* note 55, at 4 (noting that men as political beings “can experience meaningfulness only because they can talk with and make sense to each other and to themselves”).

⁵⁸ See ARENDT, THE HUMAN CONDITION, *supra* note 55, at 3-4.

⁵⁹ *Id.* at 3.

A. Arendt on Political Action and Deliberation

A question that preoccupies Arendt throughout her work is, what does it mean to belong to a political community? Following the Greek philosophers, Arendt asserts that the human capacities of action and speech are the fundamental building blocks of the political sphere.⁶⁰ Arendt defines action by distinguishing it from two other types of human activity, namely labor, which satisfies basic biological needs such as nutrition and reproduction,⁶¹ and work, which produces artificial objects that can last longer than human beings themselves.⁶² Action, by contrast, is a superior activity of public interaction between people.⁶³ Whereas labor and work take place within the private (or social) sphere, action can only take place within the public (or political) sphere.⁶⁴

Closely related to action is speech, understood as a means of expression and persuasion through words.⁶⁵ For Arendt, political life is about free discussion and debate among people, through which they define joint goals and strive to achieve them.⁶⁶ Such public deliberation allows people to see the world from other people's perspectives, and thereby to better understand it.⁶⁷ According to Arendt, this is "the political insight par excellence."⁶⁸

At the same time that action and speech constitute political life, they are also indispensable for the realization of our personality.⁶⁹ "In acting and speaking," Arendt explains, "men show who they are, reveal actively their unique personal identities and thus make their appearance in the human world."⁷⁰ Whereas the

⁶⁰ ARENDT, THE HUMAN CONDITION, *supra* note 55, at 24-26. See also *id.* at 198 ("[T]he political realm rises directly out of acting together, the 'sharing of words and deeds.'").

⁶¹ *Id.* at 7, 79-109.

⁶² *Id.* at 7, 136-175.

⁶³ *Id.* at 7, 175-181.

⁶⁴ *Id.*

⁶⁵ *Id.* at 26. With respect to the relationship between action and speech, Arendt notes that already in pre-Socratic thought, speech and action were considered to belong together, and that "this originally meant not only that most political action ... is indeed transacted in words, but more fundamentally that finding the right words at the right moment, quite apart from the information or communication they may convey, is action." *Id.* at 25-26.

⁶⁶ See Shmuel Lederman, *Agonism and Deliberation in Arendt*, 21 CONSTELLATIONS 327, 327 (2014).

⁶⁷ Hannah Arendt, *Socrates*, in THE PROMISE OF POLITICS 5, 18 (Jerome Kohn ed., 2005). See also Hannah Arendt, *Introduction Into Politics*, in THE PROMISE OF POLITICS 93, 128 (Jerome Kohn ed., 2005) (noting that the world shows itself differently to different persons and is "comprehensible only to the extent that many people can talk *about* it and exchange their opinions and perspectives with one another, over against one another").

⁶⁸ Arendt, *Socrates*, *supra* note 67, at 18.

⁶⁹ According to Arendt, personality and politics are closely interrelated. In her view, "personality is anything but a private affair." See Hannah Arendt, *Karl Jaspers: A Laudatio*, in MEN IN DARK TIMES 71, 72 (Clara & Richard Winston trans., 1958) [hereinafter Arendt, *Karl Jaspers*].

⁷⁰ ARENDT, THE HUMAN CONDITION, *supra* note 55, at 179

products of labor and work disclose *what* a person does, action and speech show *who* she is.⁷¹ In Arendt's view, "one's creations do not fully embody who one is."⁷² Accordingly, "dignity ... pertains to a man insofar as he is more than everything he does or creates."⁷³

Under the Arendtian distinction between who and what—between, on the one hand, political action and speech that realize one's personality, and, on the other hand, pre-political labor and work that are constrained by necessity and utility—participation in big data-based policymaking appears to be located too close to the non-political side of the spectrum. As noted above, the big data that policymakers use is typically produced by people for non-policy-oriented purposes within the course of mundane activities such as commuting, shopping, and communicating with family and friends.⁷⁴ In other words, such data is a creation of the private sphere that is reused for public purposes. This transplantation, however, does not attach to big data the superior virtues that Arendt attributes to real political action. Moreover, even when big data is created for policymaking purposes in the first place, it is usually technical and does not allow for genuine expression of complex thoughts and opinions.⁷⁵

Thus, whether or not big data is policy-oriented, its production and analysis can hardly reveal the "unique personal identities" of its contributors,⁷⁶ nor can they establish a meaningful public discussion that enables the participants to see reality from multiple perspectives and to find creative solutions to collective problems.⁷⁷ To use Julia Kristeva's words, big data does not have the "*energeia* that transcends deeds and [activities]," and which resists "any attempt at reification or objectification."⁷⁸ Quite to the contrary, big data reduces people's needs, experiences, and thoughts to the objectively measurable digital traces that they leave. In the world of big data, a person is not "more than everything he does or creates,"⁷⁹ but, rather, precisely what he does or creates (provided that it has a digital aspect). Hence, governmental big data analysis seems to represent a "modern prejudice," according to which "only the 'objective work,' separate from the person, belongs to the public [sphere]," while the person herself—her genuine

⁷¹ See TREVOR TCHIR, HANNAH ARENDT'S THEORY OF POLITICAL ACTION: DAIMONIC DISCLOSURE OF THE 'WHO' 127 (2017). See also JULIA KRISTEVA, HANNAH ARENDT: LIFE IS A NARRATIVE 56 (Frank Collins trans., 2001) (noting that Arendt's "who"—as opposed to "that which"—is inevitably political).

⁷² TCHIR, *supra* note 71, at 35.

⁷³ Arendt, *Karl Jaspers*, *supra* note 69, at 72.

⁷⁴ See *supra* Part I.A.

⁷⁵ See *supra* Part I.A.

⁷⁶ See *supra* note 70 and accompanying text.

⁷⁷ See *supra* notes 66-67 and accompanying text.

⁷⁸ See KRISTEVA, *supra* note 71, at 58.

⁷⁹ See *supra* note 73 and accompanying text.

thoughts, views, and feelings—remains in the shade.⁸⁰

From an Arendtian point of view, then, contemporary methods of contribution to governmental decision-making through big data either fail completely to meet the minimum standards of political participation properly so called, or, at best, represent a mild form of political participation that is far from the ideal model of such participation. While it is true that this ideal model may be impractical—even Arendt herself does not seem to believe that a comprehensive public deliberation à la Greek *polis* could take place in modern mass societies⁸¹—it should nevertheless be understood to represent a goal to which contemporary democracies, including those that use big data for policymaking purposes, should constantly aspire.⁸²

B. Arendt on Political Exclusion and the “Right to Have Rights”

As noted above, for Arendt the value of public deliberation stems from the fact that it allows people to exchange different opinions that emerge from different positions.⁸³ In other words, meaningful political participation depends on the human condition of plurality.⁸⁴ However, notwithstanding the essential differences between people, in the public sphere everyone must be considered equal.⁸⁵ This equality is an artifact of political organization: “[w]e are not born equal; we become equal as members of a group on the strength of our decision to guarantee ourselves mutually equal rights.”⁸⁶

Not all people, however, enjoy the equal opportunities offered by the public sphere. In *The Origins of Totalitarianism* Arendt discusses the situation of the paradigmatic outsider of her time—the stateless person.⁸⁷ She notes that during the first half of the twentieth century, millions of people who were members of national minorities in their countries were expelled from their homelands and lost their citizenship without being able to acquire a new one. These people became stateless,

⁸⁰ Arendt, *Karl Jaspers*, *supra* note 69, at 72.

⁸¹ See ARENDT, THE HUMAN CONDITION, *supra* note 55, at 43 (“Large numbers of people, crowded together, develop an almost irresistible inclination toward despotism, be this the despotism of a person or of majority rule.”).

⁸² As Monika Bokiniec explains, the ideal of democratic participation presented by Arendt “is not to be treated either as utopia, or as a concrete political project.” Instead, it is “an idea to direct us towards reformation or complementation of a system which failed to be democracy: a system, in which most citizens are not interested in participating in politics, and professional politicians are not interested in them participating...” See Monika Bokiniec *Is Polis the Answer? Hannah Arendt on Democracy*, 17 COACTIVITY: PHILOSOPHY, COMMUNICATION 76, 80 (2009).

⁸³ See *supra* note 67 and accompanying text.

⁸⁴ ARENDT, THE HUMAN CONDITION, *supra* note 55, at 7, 175-6, 179.

⁸⁵ *Id.* at 32-33. According to Arendt, the human condition of plurality means “living as a distinct and unique being among equals.” *See id.* at 178 (emphasis added).

⁸⁶ ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 301. *See also* Leora Bilsky, *Citizenship as Mask: Between the Imposter and the Refugee*, 15 CONSTELLATIONS 72, 74 (2008) (noting that according to Arendt, political equality is dependent upon and meaningful only within the confines of a polity).

⁸⁷ *See* ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 276-302.

lawless, and, consequently, rightless.⁸⁸ According to Arendt, this reality revealed the main paradox underlying the idea of universal human rights: precisely when they were most needed, when people who had lost everything but their humanity could rely on nothing but their “inalienable” rights as human beings, human rights proved to be an abstract, unimplementable ideal.⁸⁹

Arendt responds to the perplexities of universal human rights by introducing the concept of the ‘right to have rights.’ She defines this right as a right “to live in a framework where one is judged by one’s actions and opinions,”⁹⁰ that is, to have “a place in the world which makes opinions significant and actions effective.”⁹¹ For Arendt, the right to action is no less important than the right to be left alone, and the right to opinion is no less important than the right to think whatever one pleases.⁹² This perception of rights turns away from the liberal emphasis on negative individual freedom and focuses instead on active engagement in public life.⁹³ The right to have rights may thus be understood as a right to politics.⁹⁴ It is the right of every person to belong to a political community in which she is recognized as an equal potential partner to public deliberation over collective affairs.⁹⁵

Legal scholars and social scientists have often invoked Arendt’s concept of the right to have rights to explicate the plight of refugees and other people whose formal legal status is contested.⁹⁶ Arguably, however, this concept can also shed light on the quandary of those who are formally considered equal citizens, yet in practice

⁸⁸ *Id.* at 286-296.

⁸⁹ *Id.* at 290-300.

⁹⁰ *Id.* at 296-297.

⁹¹ *Id.* at 296. According to Seyla Benhabib, the first “right” in the “the right to have rights” refers to a moral claim to membership in an organized political community, whereas the latter refers to the legal rights of the members of such a community. See SEYLA BENHABIB, THE RIGHTS OF OTHERS: ALIENS, RESIDENTS AND CITIZENS 56-57 (2004).

⁹² ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 296.

⁹³ As noted by Jeffrey Isaac, “Arendt was a theorist of ‘positive’ rather than ‘negative’ liberty, one for whom problems of political participation and civic agency occupied center stage, and for whom the juridical strategies for limiting state power typically favored by liberals were not central.” See Jeffrey C. Isaac, *A New Guarantee on Earth: Hannah Arendt on Human Dignity and the Politics of Human Rights*, 90 AM. POL. SCI. REV. 61, 61 (1996).

⁹⁴ See James D. Ingram, What Is a “Right to Have Rights”? Three Images of the Politics of Human Rights, 102 AM. POL. SCI. REV. 401, 408-413 (2008).

⁹⁵ See *id.* at 410. See also ETIENNE BALIBAR, MASSES, CLASSES, IDEAS 212 (Swenson trans., 1994) (referring to the “universal right to political activity and recognition for every individual, in all the domains in which the problem of collectively organizing possession, power, and knowledge is posed”).

⁹⁶ See, e.g., AYTEN GÜNDODGU, RIGHTLESSNESS IN AN AGE OF RIGHTS: HANNAH ARENDT AND THE CONTEMPORARY STRUGGLES OF MIGRANTS (2015); Chenchen Zhang, *Between Postnationality and Postcoloniality: Human Rights and the Rights of Non-Citizens in a ‘Cosmopolitan Europe,’* in DECOLONIZING ENLIGHTENMENT: TRANSNATIONAL JUSTICE, HUMAN RIGHTS AND DEMOCRACY IN A POSTCOLONIAL WORLD 243 (Nikita Dhawan ed., 2014).

are excluded from various aspects of public life that are controlled by big data. While it is true that exclusion from big data does not amount to losing one's civil rights altogether and cannot be compared to statelessness or refugeehood, nothing in Arendt's account of the right to have rights suggests that it could only be applied to these radical situations. In fact, Arendt herself recognizes that the problem of exclusion from public life exists on a scale. Thus, for example, she observes that under the post-World War I international legal order, members of national minorities who were not expelled from their country and whose citizenship was not revoked were nonetheless "half stateless": they enjoyed elementary rights such as the right to life and residence, yet other rights, especially their collective cultural rights, remained in jeopardy.⁹⁷ In the same vein, it is possible to argue that while exclusion from big data does not necessarily entail absolute rightlessness, it nevertheless places significant restrictions on people's ability to enjoy equal political rights.

People who do not contribute to big data miss important opportunities to influence the public sphere. Their health problems do not affect public health programs, their commuting patterns do not affect the planning of urban transportation, their energy consumption habits do not affect governmental energy policies, and their educational achievements do not affect public schools' curricula. While other people's practices and preferences increasingly shape public policies, theirs remain untraceable and therefore irrelevant.

As noted above, the main concern of big data insiders (or of those who speak on their behalf) is that their negative freedoms—especially their rights to privacy, personal security, and freedom of thought—might be violated.⁹⁸ By contrast, big data outsiders are deprived of the positive rights to action and speech (compromised and mild as they may be); they can do whatever they want, but nobody would notice, and they can say whatever they wish, but nobody would listen. In this sense, the plight of big data outsiders resembles that of stateless persons, whose main problem is not that they are unable to exercise some civic rights, but rather that they are denied "something much more fundamental than freedom and justice,"⁹⁹ that is, the right to politics, or the right to have rights. As Serena Parekh notes, "[c]ivic rights, for Arendt, are . . . the rights of man *within a community*" (or, for that matter, within big data), whereas the rightless (or the big data outsiders) "no longer belong to any community."¹⁰⁰

Again, this does not mean that big data exiles and stateless persons are the same in all respects. However, in terms of political participation and appearance in the public sphere, exclusion from big data in the twenty first century may turn out to

⁹⁷ See ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 276.

⁹⁸ See *supra* notes 50-53 and accompanying text.

⁹⁹ See ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 296.

¹⁰⁰ See Serena Parekh, *A Meaningful Place in the World: Hannah Arendt on the Nature of Human Rights*, 3 J. HUM. RTS 41, 45 (2004) (emphasis in the original).

be analogous to exclusion from the legal framework of the state in the twentieth century. According to Arendt, the main problem of those who had lost the right to belong to an organized political community in the previous century was that they could not regain it because, due to the establishment of many new nation-states after the First World War, “there was no longer any ‘uncivilized’ spot on earth . . .”¹⁰¹ Arguably, the same can be said about big data exiles in today’s (and tomorrow’s) globalized, datafied world. As big data-based decision-making is becoming popular in many public domains, the opportunities to influence public affairs outside big data are waning.¹⁰²

To complete the analogy between stateless persons and big data outsiders, it is worth mentioning Arendt’s discussion of the dramatic mask and its metaphoric role in the political sphere.¹⁰³ In the ancient Greek theater, the dramatic mask had a double function: it hid the face of the actor, but in such a way that his voice could pass through and be heard.¹⁰⁴ According to Arendt, it is for this reason that the Latin word for mask—*persona*¹⁰⁵—is used in the legal language to denote the (artificial) legal status of a (natural) human being who belongs to an organized political community.¹⁰⁶ The idea of a “legal personality,” Arendt explains, is that when people act in the public realm, the law assigns to them a formal, equal role of right-and-duty bearers, while also allowing each of theirs authentic voices to be heard.¹⁰⁷ Put differently, the mask of the legal personality facilitates political participation and realization of human rights by covering the faces of people, that is, the attributes with which they are born such as ethnicity or sex, so that these characteristics make

¹⁰¹ ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 297.

¹⁰² Jonas Lerman similarly observes that “politicians and governments may come to rely on big data to such a degree that exclusion from data flows leads to exclusion from civic and political life—a barrier to full citizenship.” See Lerman, *supra* note 54, at 59.

¹⁰³ For another juxtaposition of the right to have rights and the metaphor of the mask, see Bilsky, *supra* note 86 (discussing Israeli citizenship law in light of Arendt’s theory).

¹⁰⁴ See Hannah Arendt, ON REVOLUTION 106 ([1963] 1990) [hereinafter Arendt, On Revolution]. See also Bilsky, *supra* note 86, at 73.

¹⁰⁵ The word *persona* is etymologically connected to the idea of sound (*sona*) projecting through (*per*) the structure of the mask. See David Marshall & Kim Barbour, *Making Intellectual Room for Persona Studies: A New Consciousness and a Shifted Perspective*, 1 PERSONA STUDIES 1, 2 (2015) (tracing the etymological origins of the Latin word “*persona*”).

¹⁰⁶ See Arendt, ON REVOLUTION, at 106-107.

¹⁰⁷ *Id.* at 107. Cf. NORMA CLAIR MORUZZI, SPEAKING THROUGH THE MASK: HANNAH ARENDT AND THE POLITICS OF SOCIAL IDENTITY 26 (2000) (contending that the metaphor of the mask is intended to propose that political action is always an assertion by a compromised, constructed identity). In addition to the legal-political context, the idea that a metaphoric mask that covers one’s face can help her sound her authentic voice and participate in public life has also been examined in various psychological and artistic works, notably in Ingmar Bergman’s 1966 film *Persona*, which presents an encounter between an actress that has ceased to speak and a caregiver that represents many aspects of the actor’s personality. See, e.g., Marshall & Barbour, *supra* note 105, at 5 (noting that the *persona* is constructed in Bergman’s film as a way to negotiate one’s social role and position).

no difference when they speak and act in the public sphere.¹⁰⁸ The public sphere, in turn, serves as a “space of appearance,”¹⁰⁹ a public stage that confers dignity and importance upon the people that appear on it.¹¹⁰ Stateless persons, however, cannot enjoy the advantages of the political mask. Deprived of the legal status of citizenship, they remain entirely exposed, caught in the “abstract nakedness of being nothing but human”¹¹¹ and treated according to their physical characteristics rather than their actions and opinions.

Big data, just like citizenship and perhaps even more so, can arguably function as an equalizing mask that hides people’s faces in a way that allows them to sound their authentic voice. The digital environment can potentially free people from identity attributes such as sex, race, and social class, which are hard to conceal in the physical world.¹¹² It can thus provide an emancipatory space of appearance where the political influence of people is determined by their own actions and not by the characteristics that define them at birth.¹¹³ Admittedly, in many cases anonymized digital data can be de-anonymized,¹¹⁴ and even when data remains anonymized, it itself or the algorithms applied to it might reproduce real-world

¹⁰⁸ See Bilsky, *supra* note 86, at 74-75. See also GEORGE KATEB, HANNAH ARENDT: POLITICS, CONSCIENCE, EVIL 10 (1983) (“Arendt presents the political actor as one who hides much in order to reveal more. He wears a mask. But the mask in the ancient theater hid the face yet allowed the actor’s true voice to come through . . . it is the highest responsibility of the citizen to protect his mask so that in the artificial composure of his appearance the truth of his words may sound.”). Cited in Bilsky, *supra* note 86, at 92, fn 6.

¹⁰⁹ See ARENDT, THE HUMAN CONDITION, *supra* note 55, at 199-207 (describing the political realm as a “space of appearance between acting and speaking men”).

¹¹⁰ See Margaret Canovan, *Politics as Culture: Hannah Arendt and the Public Realm*, in HANNAH ARENDT: CRITICAL ESSAYS, 179, 180 (Lewis P. Hinchman & Sandra K. Hinchman eds., 1994) (discussing Arendt’s conception of the public realm).

¹¹¹ ARENDT, THE ORIGINS OF TOTALITARIANISM, *supra* note 57, at 300.

¹¹² For an illuminating discussion of the ability to disrupt racial identities in cyberspace, either through anonymity or through pseudonymity (cyber-passing), see generally, Jerry Kang, *Cyber-Race*, 113 HARV. L. REV. 1131 (2000). See also Marcy Peek, *Passing Beyond Identity on the Internet: Espionage and Counterespionage in the Internet Age*, 28 VT. L. REV. 91, 93 (2003-2004) (arguing that by controlling and managing the information that they provide online, people can construct their preferred digital identity and thereby counteract discriminatory online profiling).

¹¹³ See *supra* note 90 and accompanying text. See also BENHABIB, *supra* note 91, at 59 (“The right to have rights can be realized only in a political community in which we are judged not through the characteristics which define us at birth, but through our actions and opinions, by what we do and think.”).

¹¹⁴ Computer scientists have shown that even when personally identifiable information is removed from a database, it is possible to reidentify data subjects by combining apparently non-personally identifiable information. See, e.g., Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, PROC. 2008 IEEE SYMP. ON SECURITY & PRIVACY 111; Latanya Sweeney, *Simple Demographics in the U.S. Population Uniquely*, LABORATORY FOR INT’L DATA PRIVACY, WORKING PAPER LIDAP-WP4, 2000. See also Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1707-1716 (2009) (describing the measures that database managers employ to eliminate privacy risks).

identity-based biases.¹¹⁵ However, notwithstanding these misalignments, those who can get hold of the big data mask are likely to have many more opportunities to affect public affairs than those who have no access to this mask or no ability to use it. Like their stateless twentieth century counterparts who were thrown out of the nation-state system, the twenty-first century's big data outsiders may find themselves at the margins of an emerging global political order that is increasingly shaped by big data analysis.

III. BIG DATA-BASED POLICYMAKING IN PRACTICE

A. Healthcare

Healthcare services in the United States suffer from a serious discrepancy: U.S. health expenditures per capita are the highest in the world,¹¹⁶ yet life expectancy in this country is lower than in most other developed countries.¹¹⁷ In view of this reality, improving public health and reducing health expenditures is a major policy goal for the federal government.¹¹⁸ In order to promote this goal, the Department of Health and Human Services and its agencies have recently adopted several big data projects intended to enhance the efficiency of disease prevention and treatment.¹¹⁹

¹¹⁵ See, e.g., Beth E. Kolko, Lisa Nakamura, & Gilbert B. Rodman, *Race in Cyberspace: An Introduction*, in RACE IN CYBERSPACE 1, 4-5 (Beth E. Kolko, Lisa Nakamura, & Gilbert B. Rodman eds., 2000) (“[N]either the invisibility nor the mutability of online identity make it possible for you to escape your ‘real world’ identity completely . . . all of us who spend time online are already shaped by the ways in which race matters offline.”); Baracas & Selbst, *supra* note 53, at 677-693 (describing how data mining can reflect discriminatory real world practices).

¹¹⁶ See ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD), HEALTH AT A GLANCE 2015: OECD INDICATORS 164-167 (2015) (showing that health expenditures in the U.S. are the highest among OECD countries both per capita and in terms of their share of the GDP).

¹¹⁷ *Id.* at 47 (showing that life expectancy in the U.S. is lower than the average of OECD countries).

See also LANE KENWORTHY, AMERICA’S INEFFICIENT HEALTH-CARE SYSTEM: ANOTHER LOOK (July 2011), <https://lanekenworthy.net/2011/07/10/americas-inefficient-health-care-system-another-look/> (noting that the discrepancy between life expectancy and health expenditures has been an ongoing trend in the U.S. since 1970); Ricardo Alonso-Zaldivar, *US Healthcare Spending Has Hit a New High — \$10,345 Per Person*, BUSINESS INSIDER (July 14, 2016), www.businessinsider.com/ap-new-peak-for-us-health-care-spending-10345-per-person-2016-7 (noting that national healthcare spending is expected to continue to grow in the coming years).

¹¹⁸ See, e.g., Varun Chandola et al., *Knowledge Discovery from Massive Healthcare Claims Data*, PROC. 19TH ACM SIGKDD INT’L CONF. 1312, 1312 (2013) (noting that healthcare spending is one of the key issues targeted by policymakers in the U.S.).

¹¹⁹ See, e.g., Ronald Margolis et al., *The National Institutes of Health’s Big Data to Knowledge (BD2K) Initiative: Capitalizing on Biomedical Big Data*, 21 J. AM. MED. INFORM. ASSOC. 957 (2014) (describing the National Institutes of Health BD2K initiative, which funds big data research and development projects); MCKINSEY & COMPANY, THE BIG DATA REVOLUTION IN HEALTHCARE: ACCELERATING VALUE AND INNOVATION 6 (2013), <http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care> (discussing the government’s efforts to facilitate big data analysis in

A brief description of three such projects may demonstrate this trend.

In 2013, the Centers for Disease Control and Prevention (CDC) launched the Epidemic Prediction Initiative, which aims to develop new methods to predict the outbreak and spread of infectious diseases.¹²⁰ While the CDC has been tracking influenza activity for many years now, its traditional surveillance system relies exclusively on clinical data provided by medical institutions, and therefore lags behind real-time flu activity.¹²¹ By contrast, the Epidemic Prediction Initiative supports the instantaneous mining of flu-related data drawn from social media and internet search engines.¹²² The combination of clinical and media data should enable the CDC to forecast, rather than monitor, the activity of influenza.¹²³ In the next phase, similar prediction methods are planned to be applied to other infectious diseases such as dengue fever and Ebola.¹²⁴ According to the CDC, the ultimate purpose of this project is to allow health officials to be more proactive in their prevention and mitigation strategies.¹²⁵

In 2014, the National Institutes of Health (NIH) launched the Big Data to Knowledge (BD2K) Initiative, whose purpose is to promote big data-based biomedical research in the U.S.¹²⁶ BD2K focuses on building large-scale biomedical datasets and developing relevant analysis methods. Using these resources and methods, BD2K centers across the country have already been involved in efforts to improve human understanding of, *inter alia*, the factors responsible for brain diseases, the dynamics of cardiovascular proteins, and the ways to optimize mobility in individuals with movement disabilities.¹²⁷

In 2015, the Obama administration announced the Precision Medicine Initiative.¹²⁸ The goal of this initiative is to improve the ability of healthcare providers to customize disease prevention and treatment strategies to the specific

the health sector); EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA ACROSS THE FEDERAL GOVERNMENT, *supra* note 3, at 5-6 (March 2012) (listing governmental big data initiatives in the health sector).

¹²⁰ For an overview of the Epidemic Prediction Initiative, see the program's website at predict.phiresearchlab.org/about.

¹²¹ See CENTERS FOR DISEASE CONTROL AND PREVENTION, OVERVIEW OF INFLUENZA SURVEILLANCE IN THE UNITED STATES (Oct. 13, 2016), available at www.cdc.gov/flu/pdf/weekly/overview-update.pdf (describing the current CDC flu surveillance system, which relies on data from state, local, and territorial medical institutions).

¹²² See Centers for Disease Control and Prevention, Flu Activity Forecasting Website Launched (Jan. 19, 2016), www.cdc.gov/flu/news/flu-forecast-website-launched.htm.

¹²³ See *id.*

¹²⁴ See *id.*

¹²⁵ See *id.*

¹²⁶ See Margolis et al., *supra* note 119 and accompanying text (presenting the BD2K Initiative).

¹²⁷ For an overview of these projects, see the BD2K Centers website at <https://datascience.nih.gov/bd2k/funded-programs/centers>.

¹²⁸ See Sara Reardon, *Obama to Seek \$215 Million for Precision-Medicine Plan*, NATURE (Jan. 30, 2015).

characteristics of individual patients.¹²⁹ As a first step toward this goal, the Department of Health and Human Services is working to build a national cohort of at least one million volunteers whose genetic, physiological, and behavioral data will be integrated into a massive database that will be made available to researchers and provide a platform for precision medicine analyses.¹³⁰

In terms of data sources, the three initiatives described above rely mainly on electronic health records (EHRs) collected from hospitals, medical clinics, public health laboratories, and medical insurance companies, and containing such data as imaging, laboratory test results, genetic information, diagnoses, and prescribed medications.¹³¹ The initiatives also rely on social media exchanges and internet searches,¹³² as well as on data produced directly by patients who self-report their symptoms, pain scales, smoking habits, etc., or who use mobile sensors that monitor their blood pressure, heart rate, glucose level, etc.¹³³ These sources are not unique to the initiatives discussed here; they are the building blocks of many (probably most) current and prospective healthcare big data projects.¹³⁴

¹²⁹ See Precision Medicine Initiative Working Group, The Precision Medicine Initiative Cohort Program: Building a Research Foundation for 21st Century Medicine 6 (Sept. 17, 2015), www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf.

¹³⁰ See *id.*

¹³¹ See, e.g., CENTERS FOR DISEASE CONTROL AND PREVENTION, *supra* note 122 (mentioning the role of EHRs in the Epidemic Prediction Initiative); Derrek P. Hibar et al., *Cortical Abnormalities in Bipolar Disorder*, 23 MOLECULAR PSYCHIATRY 932, 933 (2017) (describing a research conducted by a BD2K center, which analyzed MRIs and other EHR data to shed light on the physiology of bipolar disorder); PRECISION MEDICINE INITIATIVE WORKING GROUP, *supra* note 129, at 21 (noting that “A substantial source of health data for the [Precision Medicine Initiative] cohort should derive from EHRs”).

¹³² See, e.g., CENTERS FOR DISEASE CONTROL AND PREVENTION, *supra* note 122 (explaining the role of social media and internet searches in the Epidemic Prediction Initiative).

¹³³ See, e.g., PRECISION MEDICINE INITIATIVE WORKING GROUP, *supra* note 129, at 21 (noting that in addition to EHRs, the Precision Medicine Initiative cohort “will also include self-report data and the ability to collect data from a number of other sources, such as sensors and mobile devices”); Santosh Kumar et al., *Center of excellence for mobile sensor data-to-knowledge* (MD2K), 22 J AM. MED. INFORM. ASSOC. 1137 (2015) (describing the mission of a BD2K center that develops tools to streamline the collection and analysis of data generated by mobile health sensors).

¹³⁴ See, e.g., Tao-Huang et al., *Promises and Challenges of Big Data Computing in Health Sciences*, 2 BIG DATA RESEARCH 2, 4-5 (2015) (noting that mobile phones, social media, and data from hospitals are the main sources of medical big data); Meredith A. Barrett et al., *Big Data and Disease Prevention: From Quantified Self to Quantified Communities*, 1(3) BIG DATA 168 (Sept. 2013) (discussing the contribution of big data drawn from EHRs and mobile health sensors to the prevention of chronic diseases); Christian R. Macedonia et al., *Advanced Research and Data Methods in Women’s Health: Big Data Analytics, Adaptive Studies, and the Road Ahead*, 129 OBSTETRICS & GYNECOLOGY 249 (2017) (highlighting the importance of large-scale EHR analyses in the field of gynecology); Sean D. Young, “*Big Data*” Approach to HIV Epidemiology and Prevention, 70 PREV. MED. 17 (2015) (examining ways to use social media and mobile technology data to address the HIV epidemic); David W. Bates et al., *Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients*, 33 HEALTH AFFAIRS 1123, 1123 (2014)

As we can see, most big data analysis in the healthcare sector relies on the secondary use of non-policy-oriented data that was originally generated for other purposes. Medical institutions store patients' information in EHRs in order to offer them more efficient treatment over time and to prevent the loss of information.¹³⁵ While the collection and storage of such information as well as its reuse by government agencies usually requires patients' consent,¹³⁶ this consent does not mean that the patient undertook medical examinations with the intention of affecting public policy or that she actually contemplated possible policy implications. The same is true of spontaneous searches for medical information on the internet and the sharing of such information in social networks—people who engage in such activities do not usually intend that it will serve for policymaking purposes. Finally, when it comes to self-monitoring and direct reporting by citizens, the data analyzed by health authorities is policy-oriented from the outset, yet it only includes technical information (reporters' smoking and drinking habits, blood pressure, etc.), without any reference to related policy issues.

Hence, whatever the source of medical big data, its deployment for policymaking purposes can offer data contributors no more than a mild form of political participation. The collection and analysis process ensures that policymakers take into account or at least be aware of contributors' "objective" health conditions, yet it does not give the latter an opportunity to present their views and insights about health policy questions. It assigns weight to their physiological conditions and needs, but not to their thoughts and ideas, and thus makes no room for their unique personalities to appear in the public sphere and interact with each other. Put differently, contemporary public engagement in health policymaking through big data does not seem to meet Arendt's basic requirements of meaningful political participation.

In any event, the situation of those who contribute technical information to medical databases appears to be better than the situation of those who cannot even enter the big data loop. Although the data sources used by health authorities are quite diverse, many U.S. citizens are left outside them. To begin with, millions of Americans cannot afford the costs of visiting hospitals and medical clinics, and therefore their medical data does not appear in EHRs. In 2015, 9.1 percent of the

(discussing EHRs' big data implications); Munmun De Choudhury et al., *Predicting Postpartum Changes in Emotion and Behavior via Social Media*, 2013 ACM Annual Conference on Human Factors in Computing Systems 3267 (2013) (suggesting that social media data can be used to identify early signs of mental illness).

¹³⁵ See, e.g., David Blumenthal & Marilyn Tavenner, *The "Meaningful Use" Regulation for Electronic Health Records*, 363 N. ENGL. J. MED. 501, 503 (2010) (noting that EHRs can improve the safety, quality, and efficiency of care and help clinicians avoid preventable errors).

¹³⁶ See, e.g., Peter B. Jensen et al., *Mining Electronic Health Records: Towards Better Research Applications and Clinical Care*, 13 NATURE REVIEWS: GENETICS 395, 402 (2012) (noting that patients' agreement to the secondary use of EHR data is commonly obtained through an opt-in model as part of the standard treatment consent form).

U.S. population, or 29 million people, had no health insurance.¹³⁷ Though undoubtedly high, this number nonetheless represents a significant decrease in the size of the uninsured population following the enactment of the Patient Protection and Affordable Care Act in 2010 and the consequent expansion of Medicaid eligibility in 2014.¹³⁸ However, if the current administration's plan to repeal major parts of this Act is ultimately realized, the number of uninsured persons is likely to rise again.¹³⁹ Furthermore, even among people who do have health insurance, representation in EHRs is not equal. One reason for that is that large, technologically advanced medical institutions located in economically developed areas make greater use of EHRs than smaller, peripheral institutions that usually serve poorer populations.¹⁴⁰

Arguably, those who cannot afford to see a physician and those who visit hospitals and clinics that do not use EHRs could still contribute to some medical big data projects by independently reporting their physiological and behavioral conditions or by using automatic mobile health applications that monitor and report such information directly to the relevant databases. Yet, although this type of participation does not depend on one's ability to pay for institutional healthcare services, it does require other resources such as time, trust, and a certain socio-psychological orientation. For example, in order to participate in the national cohort created under the Precision Medicine Initiative, people are required to monitor and share their health data and to provide biospecimens over a period of at least ten years.¹⁴¹ Unfortunately, studies on civic engagement show that people from a disadvantaged social background are less likely to be committed to such a long-term voluntary project.¹⁴²

Finally, some medical big data projects rely on social media and internet

¹³⁷ See U.S. Census Bureau, *Health Insurance Coverage in the United States: 2015*, page 1, CURRENT POPULATION REPORTS, P60-257(RV) (Sept. 2016).

¹³⁸ See National Center for Health Statistics, Long-term Trends in Health Insurance Coverage: Estimates from the National Health Interview Survey, 1968–2016, page 4 (July 2017) (showing that the number of uninsured persons under the age of 65 dropped from 48.3 million in 2010 to 28 million in 2016). See also U.S. Census Bureau, *supra* note 137, at 2 (attributing this drop to the enactment of the Affordable Care Act).

¹³⁹ Repealing and replacing Obamacare was a central pledge of Trump's 2016 campaign, and his administration has invested many efforts to achieve this goal. While Trump's suggested reform has so far failed to achieve sufficient support in Congress, it is still too early to declare it dead. See, e.g., Editorial, *Obamacare Saved, But Just Barely*, N.Y. TIMES, July 29, 2017, at A24.

¹⁴⁰ See, e.g., BLUMENTHAL & TAVENNER, *supra* note 135, at 501 (noting that despite a broad consensus regarding the potential value of EHRs, they are still unavailable in many medical institutions).

¹⁴¹ See PRECISION MEDICINE INITIATIVE WORKING GROUP, *supra* note 129, at 21.

¹⁴² See, e.g., J. Foster-Bay, *Do Race, Ethnicity, Citizenship and Socio-economic Status Determine Civic Engagement?* (Ctr. for Information & Research on Civic Learning & Engagement, Working Paper No. 62, 2008), <http://files.eric.ed.gov/fulltext/ED505266.pdf> (finding that the propensity to volunteer and to remain civically engaged over time is confounded by income, education, race, ethnicity, and citizenship status).

searches to analyze health-related practices and trends. Here, too, the conditions and needs of members of socioeconomically disadvantaged groups are likely to be underrepresented. As noted above, poorer and less educated people do not use the internet as much as wealthier people, and even when they do, they may not use it to obtain or share information that could be relevant for public policymaking, such as medical information.¹⁴³

Hence, whether medical big data is extracted from institutional EHRs, self-monitoring, or internet activity, members of disempowered groups are less likely to participate in it. Consequently, the health needs of poorer populations might have a smaller impact on the determination of public health policy than the needs of other groups. This conclusion is particularly troubling in view of the fact that poorer populations are generally less healthy and have a shorter life expectancy than other populations. This is so not only because poorer people cannot afford adequate healthcare, but also due to other factors such as greater exposure to environmental hazards, more frequent workplace and home injuries, and higher rates of smoking and drinking,¹⁴⁴ which increase the prevalence of chronic diseases and physical disabilities.¹⁴⁵ This means that those who are underrepresented in medical big data are precisely the ones who have the greatest interest in its policy outcomes.

One may question the relevance and legitimacy of this discussion and argue that even if socioeconomically disadvantaged people do not have equal opportunities to affect big data-based health policymaking, this is probably the last thing that concerns them. A person whose medical data does not appear in EHRs because he does not have health insurance cares about receiving immediate healthcare services, not about the negligible policy implications of his absence from medical databases. Moreover, it could be argued that focusing on the long-term policy implications of the exclusion of disadvantaged populations from big data might turn public attention away from the immediate hardships experienced by these populations. This criticism, however, fails to acknowledge that an effective response to the hardships of underserved populations must include not only immediate remedies, but also long-term policy change, which cannot take place unless these populations participate in the shaping of relevant policies, *inter alia* by producing big data.¹⁴⁶

¹⁴³ See *supra* notes 36-40 and accompanying text.

¹⁴⁴ See, e.g., Fred C. Pampel et al., *Socioeconomic Disparities in Health Behaviors*, 36 ANNU. REV. SOCIOOL. 349, 350 (2010) (reviewing various explanations for unhealthy behaviors among low-status socioeconomic groups).

¹⁴⁵ See Centers for Disease Control and Prevention, *Health Disparities and Inequalities Report — United States, 2013*, 62 MORBIDITY AND MORTALITY WEEKLY REPORT (MMWR), Suppl. 3, 15 (2013) (observing that lower socioeconomic position is associated with higher morbidity and mortality and mentioning behavioral and environmental factors that account for this correlation).

¹⁴⁶ See, e.g., Ursula E. Bauer & Marcus Plescia, Addressing Disparities in the Health of American Indian and Alaska Native People: The Importance of Improved Public Health Data, 104 (Suppl. 3) AM. J. PUB. HEALTH 255 (2014) (noting that data gaps pose an obstacle to creating health policies that adequately address the particular health needs of the Native American population, such as extremely high rates of diabetes and chronic liver disease).

B. Urban Planning

As the world's population rapidly grows and urbanizes,¹⁴⁷ adopting advanced, data-based solutions to the increasing pressure on urban infrastructure becomes an imperative.¹⁴⁸ From China to Singapore to Abu Dhabi to Spain, national and local governments around the world are attempting to turn crowded cities into "smart cities" that use resources and provide services in an informed, efficient, and sustainable way.¹⁴⁹ The United States is no exception. In the last couple of years, municipal authorities across the country have launched numerous smart city projects,¹⁵⁰ often with the support of the federal government.¹⁵¹ These projects use big data analytics in order to save energy, reduce pollution, improve sanitation, facilitate the flow of traffic, and offer overall better services to the residents.¹⁵²

Smartphone applications are a major source of data for many of these projects. For example, the departments of transportation in Alabama, Florida, Georgia, Massachusetts, Oregon, and in several other states and municipalities have recently entered into agreements with the crowdsourced navigation application Waze, according to which Waze will share with them traffic and road condition reports that it collects from drivers.¹⁵³ The idea, according to Waze, is to improve the day-to-day management as well as the longer-term planning of urban transportation.¹⁵⁴ Another popular smartphone application from which local authorities extract big data is Strava Metro, which tracks users' cycling activity and aggregates the data to identify common cycling routes and times.¹⁵⁵ Municipal authorities in Portland,

¹⁴⁷ The urban population of the world has grown from 746 million in 1950 to 3.9 billion in 2014 and is expected to grow to 6.3 billion by 2050. See United Nations Department of Economic and Social Affairs, *World Urbanization Prospects: The 2014 Revision*, ST/ESA/SER.A/352, 11 (2015). The United States is among the most urbanized countries in the world, with about 265 million city dwellers in 2015, expected to grow to 350 million by 2050 (*id.* at 225).

¹⁴⁸ See *id.* at 3 (noting that "rapid and unplanned urban growth threatens sustainable development when the necessary infrastructure is not developed").

¹⁴⁹ Smart cities may be defined as "places where information technology is combined with infrastructure, architecture, everyday objects, and even our bodies to address social, economic, and environmental problems." ANTHONY M. TOWNSEND, SMART CITIES: BIG DATA, CIVIC HACKERS, AND THE QUEST FOR A NEW UTOPIA 15 (2014).

¹⁵⁰ See Executive Office of the President of the United States, Report to the President: Technology and the Future of Cities 8-27 (Feb. 2016) [hereinafter Executive Office of the President, Technology and the Future of Cities] (providing an overview of ongoing and planned smart city projects in the United States).

¹⁵¹ See *id.* at 62-68 (describing major federal government smart city initiatives, including the 2013 Department of Commerce Smart America Challenge, the 2015 Department of Transportation Smart City Challenge, and the 2015 White House Smart Cities Initiative).

¹⁵² *Id.* at 1-3 (summarizing the main goals of smart city projects).

¹⁵³ See Waze's Connected Citizens Program website, at <https://www.waze.com/ccp>.

¹⁵⁴ See Ryan Bradley, *Waze and the Traffic Panopticon*, THE NEW YORKER (June 2, 2015), www.newyorker.com/business/currency/waze-and-the-traffic-panopticon (waze aims to "improve some of [Los Angeles'] traffic and infrastructure systems . . .")

¹⁵⁵ See STRAVA METRO APPLICATION, <https://metro.strava.com> (last visited Oct. 26, 2018).

Seattle, Orlando, and other major U.S. cities use this data to accommodate their road and transportation design to cyclers' needs.¹⁵⁶

In addition to the secondary use of non-policy-oriented applications like Waze and Strava Metro, many local authorities also extract data from designated smart city mobile applications. A famous example is Street Bump, which has been developed and used by the City of Boston and is expected to also become available to other cities. This application automatically reports to the local authorities whenever a user drives over a pothole,¹⁵⁷ thereby allowing the authorities to make immediate repairs and at the same time to better plan road infrastructure and maintenance.¹⁵⁸ Another notable example is 311 applications that allow city residents to report various types of hazards and impediments. Here, too, municipal authorities use the reports both to solve immediate problems (e.g., fix a sewage leak) and to improve relevant systems in the longer term (e.g., redesign sewage infrastructure on the basis of accumulated data).¹⁵⁹

Another, relatively new, source of data for smart city analytics is personal sensors. Such sensors are currently being used in several cities in Europe, including Amsterdam, Barcelona, and Manchester. In these cities, some residents have voluntarily installed on their apartments' windows Smart Citizen Kits that collect environmental data.¹⁶⁰ The kit contains a board with several sensors that measure air composition, light intensity, noise levels, and other environmental data. This data is constantly streamed to a common platform that can be accessed by municipal authorities.¹⁶¹ Currently, it seems, no U.S. local government is systematically extracting data from personal sensor kits.¹⁶² However, several academic research groups have received governmental funding to develop such systems. For example, a group from the University of Michigan is developing the Sensors in a Shoebox Kit, which can measure environmental parameters such as air quality, humidity,

¹⁵⁶ See, e.g., Peter Walker, City Planners Tap into Wealth of Cycling Data from Strava Tracking App: Seventy-Six Cities and Regions are Using Strava Metro Data to help Assess and Shape Transport Policy, THE GUARDIAN (May 9, 2016), available at www.theguardian.com/lifeandstyle/2016/may/09/city-planners-cycling-data-strava-tracking-app.

¹⁵⁷ See City of Boston, *Street Bump* (last updated Apr. 21, 2017), <http://www.boston.gov/departments/new-urban-mechanics/street-bump>.

¹⁵⁸ See *id.*

¹⁵⁹ See, e.g., BOSTON 311 APPLICATION (Bos:311) <https://www.boston.gov/departments/innovation-and-technology/apps>; LOS ANGELES 311 APPLICATION (MyLA311) <https://myla311.lacity.org/>; and NEW YORK CITY 311 APPLICATION (NYC311) <https://www1.nyc.gov/311/index.page>.

¹⁶⁰ *Digital Social Innovation, Smart Citizen: A Kit and Platform which Engages Citizens in Solving Local Environmental Problems* (October 2016), <https://digitalsocial.eu/case-study/9/smart-citizen>.

¹⁶¹ See SMART CITIZEN ABOUT, <https://smartcitizen.me/about> (last visited Oct. 26, 2018).

¹⁶² No relevant information was displayed when the author searched on Google the terms "city/urban" and "sensor" combined with the names of the fifty most populated cities in the United States according to the U.S. Census Bureau's 2016 Population Estimates, <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk> (last visited Oct. 26, 2018).

and pedestrian traffic.¹⁶³ The kits are planned to be operated by high school students in the City of Detroit, who will be able to learn about their neighborhoods and provide relevant data to decision-makers.¹⁶⁴ Another group of researchers from New York University is developing advanced microphones that will be able to record and identify various types of urban sounds, from the noise of an air conditioner to a dog's barking.¹⁶⁵ Developers hope that these microphones will eventually be used by New York City residents to understand noise pollution in their city and to assist city officials in addressing the problem.¹⁶⁶

Finally, some municipal authorities collect data from public sensors that they themselves install and operate in public places. An example in point is the Array of Things project implemented by the City of Chicago. Under this project, the City of Chicago plans to install on traffic light poles across the city five hundred sensors (nodes) that will measure weather conditions, air quality, and pedestrian and vehicular traffic.¹⁶⁷ Ultimately, this data will help city agencies study and address critical city challenges such as improving air quality and traffic safety and preventing urban flooding.¹⁶⁸

In terms of civic engagement, all of the abovementioned smart city initiatives seem to suffer from the dual chronic disease characteristic of big data-based policymaking. First, these initiatives establish a mild form of political participation. They offer city residents representation but not voice, consideration but not deliberation. This is obviously true of smartphone applications like Waze and Strava Metro, which people use with no policy motivations in mind. However, it is also true of policy-oriented applications such as Street Bump, which provide people with the opportunity to influence city affairs directly and purposefully, but only in a technical, semi-automatic manner.

311-like applications are somewhat different in that they are more open-ended and allow citizens greater control over the information and, no less importantly, the specific message that they transfer to the authorities. Theoretically speaking, city

¹⁶³ See Ted Montgomery, *U-M School of Education, College of Engineering Partner to Win Knight Cities Challenge Award*, UNIVERSITY OF MICHIGAN NEWS (Apr. 12, 2016), available at <http://ns.umich.edu/new/releases/23700-u-m-school-of-education-college-of-engineering-partner-to-win-knight-cities-challenge-award>. See also Amy Crawford, *Detroit Imagines a Citizen-Led Smart City*, CITYLAB (May 31, 2017), available at <https://www.citylab.com/life/2017/05/detroit-imagines-a-citizen-led-smart-city/528441/>.

¹⁶⁴ See *id.*

¹⁶⁵ See Emily S. Rueb, *To Create a Quieter City, They're Recording the Sounds of New York*, N.Y. TIMES, page A19 (Nov. 7, 2016) ("[a] group of researchers from New York University and Ohio State University are training the microphones to recognize jackhammers, idling engines and street music, using technology originally developed to identify the flight calls of migrating birds").

¹⁶⁶ See *id.*

¹⁶⁷ University of Chicago Press Release, *Chicago Becomes First City to Launch Array of Things* (Aug. 29, 2016), <https://ci.uchicago.edu/press-releases/chicago-becomes-first-city-launch-array-things> ("[t]he first two nodes . . . will collect information on weather, traffic, and air quality").

¹⁶⁸ See *id.*

residents could use these applications to express their views on a variety of municipal issues and offer their personal policy advice. In practice, however, it is doubtful that local authorities currently have the technical capacity to conduct systematic large-scale analyses of such complex data. Until such capacities become operational, 311 applications are unlikely to provide a platform for meaningful political participation through big data.¹⁶⁹

Sure enough, personal sensors offer no better opportunities for people to engage in public deliberation over urban affairs. In fact, when local authorities rely on personal sensors to inform policymaking, the very particular role they assign to people is to be device carriers, demonstrating anything but respect for their capacities as rational agents. This situation seems to perfectly realize Arendt's concerns about the rise of machines and the objectification of human beings.¹⁷⁰ It also seems safe to assume that Arendt would have been uncomfortable with the widespread use of public sensors that are installed on street poles and which communicate information to policymakers, quite literally, over people's heads, giving them no role whatsoever in the policymaking process.

Second, smart city projects of the kind discussed above are likely to have an exclusionary effect on socioeconomically disadvantaged groups. A smartphone application like Waze, Strava Metro, Street Bump, or 311 can only be used by people who own a smartphone and have downloaded and know how to operate the application. Since people with limited financial resources are less likely than others to be such,¹⁷¹ lower-class neighborhoods may not equally benefit from the services that municipal authorities provide and from the policies that they adopt on the basis of data retrieved from mobile applications.

When it comes to environmental sensors the reasons for the underrepresentation of socioeconomically disadvantaged populations are somewhat different and not necessarily connected with the "digital divide." Unlike big data projects that draw information from mobile phone applications in an apparently random bottom-up manner, projects that rely on environmental sensors are more likely to be controlled from above. This is true for public sensors that local authorities directly install and operate, and it may also be true for personal sensors if the authorities allocate them to a selected group of residents. In these cases, local officials might distribute sensors in an unequal manner, awarding higher priority to identifying the environmental needs of wealthier residents, which usually have greater influence on local politics.¹⁷² For example, in Chicago's Array of Things project, the location

¹⁶⁹ See *infra* notes 210-211 and accompanying text.

¹⁷⁰ See *supra* notes 55-59 and accompanying text.

¹⁷¹ It has been observed, for example, that leisure and sport cyclists tend to use the Strava Metro application more than people who ride their bikes to work, and that people who live in Manhattan are two to three times more likely to report disturbances through New York City's 311 application than their poorer neighbors in Brooklyn, Queens, and the Bronx. See Walker, *supra* note 156; Rueb, *supra* note 165.

¹⁷² The tendency of government agents to award higher priority to the interests of more powerful

of more than a hundred nodes have already been determined. A close examination reveals that hardly any nodes are currently installed or planned to be installed in the city's poorer neighborhoods.¹⁷³ This example demonstrates that when the authorities themselves control the production of big data, they might reproduce the exclusionary patterns that often characterize traditional policymaking processes.

It has been noted above that disadvantaged populations who have a relatively small influence on big data-based health policymaking are usually the ones who have the greatest interest in public health policy. Apparently, the same is true of those excluded from big data-based urban planning.¹⁷⁴ Poorer people who live in poorer neighborhoods usually suffer more from noise, air pollution, and other annoyances, and are more dependent on publicly funded services to address these impediments.¹⁷⁵ In order to provide municipal services to those who need them the most, it is necessary to find ways to realize the inclusive potential of big data-based decision-making.

IV. IS THERE A CONSTITUTIONAL RIGHT TO MEANINGFUL AND INCLUSIVE PARTICIPATION?

The upshot of the previous Parts is that there is a significant gap between the alleged participatory potential of big data-based policymaking and its actual impact on public engagement in political decision-making. On the one hand, big data technologies allow government officials to take into account much more information relating to many more people than they ever could before, and thus to adopt policies and regulations that are more responsive to the needs of greater

groups within their constituencies has been extensively discussed in public choice literature. See, e.g., MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION ([1965] 1971) (explaining why small, well-organized interest groups have considerable influence over public decision-makers); Gary Becker, *A Theory of Competition among Pressure Groups for Political Influence*, 98 Q.J. ECON. 371 (1983) (explaining why government agents promote policies that they believe to be favored by powerful groups).

¹⁷³ See Chicago Data Portal, *Array of Things Locations*, <https://data.cityofchicago.org/Environment-Sustainable-Development/Array-of-Things-Locations/6rq2-yx28/data> (updated December 20, 2018) (providing a list of the locations of 111 nodes); *Array of Things, Node Location Map*, arrayofthings.github.io/node-locations.html (presenting the nodes on the city map); Chicago Data Portal, *Census Data*, data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2 (updated Sept. 12, 2014) (drawing on national census data to rank Chicago's neighborhoods [community areas] according to selected socioeconomic indicators).

¹⁷⁴ See, e.g., EXECUTIVE OFFICE OF THE PRESIDENT, TECHNOLOGY AND THE FUTURE OF CITIES, *supra* note 150, at 8 (noting that the “[c]hallenges faced by Americans living in cities... are often intensified for those who are poor, disabled, young, alone, or aged” and that “[t]hese same disadvantaged groups also often have the least opportunity to take direct advantage of new technologies”).

¹⁷⁵ See, e.g., Edmund Y.W. Seto et al., *Spatial distribution of traffic induced noise exposures in a US city*, 6 INT'L J. HEALTH GEOGRAPHICS 24 (2007) (finding that in San Francisco, the risk of annoyance from urban noise varies considerably between neighborhoods).

segments of the population. On the other hand, big data can only represent people's needs in a superficial and schematic manner, and it cannot offer them real voice or meaningful deliberation. In addition, big data's inclusiveness seems to end at the doorstep of socioeconomically disadvantaged populations who do not produce relevant digital data.

This Part sets out to examine whether constitutional law can play a role in reducing the gap between the promise and reality of political participation through big data. Section A suggests that the First Amendment may be understood to place upon public authorities that rely on big data analysis a duty to ensure that it does not deny people the opportunity to express and discuss their views in a meaningful way. Section B suggests that the Fourteenth Amendment may be understood to place upon the same authorities a duty to ensure that all those potentially affected by their decisions, including members of socioeconomically disadvantaged groups, have an equal opportunity to participate. Both sections discuss concrete measures that the government can take to dispose of these purported constitutional duties.

A. *Mild Participation and the First Amendment*

The U.S. Constitution does not explicitly mention the right of citizens to participate in political or public affairs. In fact, even the narrower right to vote in periodic elections, which is a core component of the right to political participation,¹⁷⁶ is only indirectly embedded in the Constitution.¹⁷⁷ It could be argued, however, that in view of its centrality to the Founding Fathers' legal and political vision, the right to political participation should nonetheless be understood to underlie American constitutional law and should inspire the interpretation of various constitutional norms.¹⁷⁸

But how exactly did the Constitution's Framers conceive of the right to political participation? What was the content and scope of this right in their vision? For the Framers, the main purpose of political participation was to ensure that the government promotes the common good of the citizenry and refrains from using its power in an abusive or arbitrary manner.¹⁷⁹ This goal apparently implied a rather narrow conception of political participation, which focused on the right to vote. As James Madison explains in *The Federalist* no. 39, "[i]t is *essential* to [the]

¹⁷⁶ See, e.g., Henry J. Steiner, *Political Participation as a Human Right*, 1 HARV. HUM. RTS. Y.B. 77, 99-100 (1988) (distinguishing between "thin" conceptions of political participation, which view elections as a "near-sufficient form of political participation," and "strong" conceptions of political participation, which argue for an ongoing right to take part in the conduct of public affairs in various non-electoral ways that supplement voting).

¹⁷⁷ See U.S. Const. amends. XV, XIX, XXIV, and XXVI (prohibiting abridgment of the right to vote on account of race, sex, or age or due to a failure to pay a poll tax).

¹⁷⁸ See *infra* notes 185-190 and accompanying text.

¹⁷⁹ See, e.g., Sunstein, *Interest Groups*, *supra* note 32, at 31 (noting that the Framers adhered to a republican conception according to which politics consists of self-rule by citizens willing to subordinate their factional interests to the general good).

government that [its power] be derived from the great body of the society, not from an inconsiderable proportion, or a favored class of it ... It is *sufficient* for such a government that the persons administering it be appointed, either directly or indirectly, by the people.”¹⁸⁰ According to Madison, the participation of a large body of voters in elections provides both a necessary and sufficient guarantee that the citizens will choose their representatives according to the latter’s ability to promote the public good—that is, according to their civic virtue—rather than on the basis of their affiliation with some factionist interests.¹⁸¹ At the same time, the periodic nature of elections ensures that representatives remain accountable to their constituencies, who will be able to replace them in case they turn against the public interest.¹⁸²

In the Framers’ vision, once they are elected, the representatives should make decisions through a deliberative process that helps them discern the common good of the society and find the best ways to pursue it. This deliberative process takes place within the representative body, without involving the public. According to the Framers, such internal deliberation may yield decisions that are “more consonant to the public good than if pronounced by the people themselves,” for the wisdom of the representatives “may best discern the true interest of their country,” and their “patriotism and love of justice” will make them “least likely to sacrifice it to temporary or partial considerations.”¹⁸³

The constitutional framework envisioned by the Framers thus appears to have an elitist bent. It assigns the task of governing to the most virtuous persons who can best promote the people’s collective interest, whereas the role of citizens is largely confined to identifying these persons and choosing them as their representatives.¹⁸⁴

¹⁸⁰ THE FEDERALIST No. 39, at 241 (J. Madison) (Clinton Rossiter ed. 1961) (emphasis in the original). Cited in Sunstein, *Beyond the Republican Revival*, *supra* note 30, at 1548, fn 41.

¹⁸¹ Madison’s idea is that in large constituencies each representative is chosen by a large number of citizens with diverse interests, and therefore there are fewer chances that “unworthy candidates” will be able to offer their supporters selective benefits. See THE FEDERALIST, *supra* note 180, No. 10, at 62 (J. Madison).

¹⁸² THE FEDERALIST, *supra* note 180, No. 39, at 241 (J. Madison) (noting that elected representatives should be “holding their offices during pleasure, for a limited period, or during good behavior”). In addition to general periodic elections, the Framers made sure to include in the Constitution other mechanisms that would constrain the power of the government, such as the separation of powers and the protection of individual liberties.

¹⁸³ THE FEDERALIST No. 10, *supra* note 180, at 60 (J. Madison). Madison further observes that “[t]he aim of every political constitution is, or ought to be, first to obtain for rulers men who possess most wisdom to discern, and most virtue to pursue the common good of the society; and in the next place, to take the most effectual precautions for keeping them virtuous whilst they continue to hold their public trust.” See THE FEDERALIST No. 57, *supra* note 180, at 377 (J. Madison). Both sources are cited in Sunstein, *Interest Groups*, *supra* note 32, at 41.

¹⁸⁴ See Frank I. Michelman, *The Supreme Court 1985 Term—Foreword: Traces of Self-Government*, 100 HARV. L. REV. 4, 20 (1986) [hereinafter Michelman, *Traces of Self-Government*] (explaining that the assertion that “there is a natural aristocracy of talent, for whom the tasks of government should, in everyone’s interest, be reserved,” may be seen as elitist).

Some scholars suggest, however, that a careful examination of the Framers' work and of the historical context in which they operated reveals that their conception of civic participation was broader than it initially seems. These scholars observe that the Framers were heavily influenced by republican ideals, which assign citizens a greater role in politics than mere participation in elections.¹⁸⁵ Classic republicanism views dialogue and discussion among the citizenry as crucial for discerning the common good as well as for inculcating civic virtue and enhancing individual self-realization.¹⁸⁶ It therefore asserts that "political participation should be active and frequent and not limited to voting."¹⁸⁷ As Cass Sunstein explains, these republican principles were not fully or coherently endorsed by the Founding Fathers.¹⁸⁸ However, they did play an important role in the constitutional framing process,¹⁸⁹ and they should and do continue to inspire contemporary constitutional doctrine and interpretation.¹⁹⁰

Most notably, the republican ideal of participatory politics finds expression in the First Amendment, which enshrines, *inter alia*, the freedom of speech and the press, the right to peaceably assemble, and the right to petition the government.¹⁹¹ By protecting these freedoms from abridgement by the federal and state

¹⁸⁵ See, e.g., Sunstein, *Interest Groups*, *supra* note 32; Sunstein, *Beyond the Republican Revival*, *supra* note 30; Bruce Ackerman, *The Storrs Lectures: Discovering the Constitution*, 93 YALE L.J. 1013 (1984); Frank Michelman, *Law's Republic*, 97 YALE L.J. 1493 (1988); Michelman, *supra* note 184. One of the purposes of the "Republican Revival" in constitutional legal scholarship was to refute the assumption that the Constitution's Framers were exclusively influenced by classic liberalism or pluralism. See Richard H. Fallon Jr., *What Is Republicanism, and Is It Worth Reviving*, 102 HARV. L. REV. 1695, 1695-96 (1989).

¹⁸⁶ See, e.g., GEOFFREY R. STONE, LOUIS MICHAEL SEIDMAN, CASS R. SUNSTEIN, MARK V. TUSHNET, & PAMELA S. KARLAN, *CONSTITUTIONAL LAW* 7 (8th edition, 2017).

¹⁸⁷ *Id.* at 7.

¹⁸⁸ See Sunstein, *Beyond the Republican Revival*, *supra* note 30, at 1558-1560.

¹⁸⁹ Sunstein explains that republican ideas were more strongly advocated by the antifederalists than by the federalists. However, important elements of republican thought can also be found in the views of the federalists. See *id.* at 1547.

¹⁹⁰ See *id.* at 1540 (noting that "[r]epublican thought played a central role in the framing period, and it offers a powerful conception of politics and of the functions of constitutionalism," and that "[t]he characteristically republican belief in deliberative democracy continues to influence... legal doctrine.") According to Sunstein, the republican ideal of deliberative democracy finds its expression in the American constitutional framework, *inter alia*, in the adoption of a federal system that provides opportunities for local self-determination at the state level, as well as in the right to a jury trial, which is valued in the republican tradition as a means of inculcating civic virtue and promoting participation (*id.* at 1562). See also Michelman, *Traces of Self-Government*, *supra* note 184, at 23 (noting that "the republican tradition of civic dialogue retains a strong, if somewhat disguised and twisted, hold on American constitutional imagination").

¹⁹¹ See, e.g., Sunstein, *Beyond the Republican Revival*, *supra* note 30, at 1550 (asserting that the requirement of deliberation embodies substantive limitations such as the protection of the freedom of speech); Cass R. Sunstein, *The First Amendment in Cyberspace*, 104 YALE L.J. 1757, 1762 (1995) ("The Madisonian model sees the right of free expression as a key part of the system of public deliberation").

governments,¹⁹² this Amendment arguably establishes the basic conditions for authentic public deliberation, which cannot take place unless people are free to express their views.¹⁹³ According to this reading, the main purpose of the Framers in adopting this amendment was to facilitate informal public deliberation that takes place in the streets, in town squares, in the press, and through lobbies and petitions, and which complements the internal deliberations that take place within official government institutions.¹⁹⁴

The first to offer a comprehensive account of the deliberative foundations of the First Amendment was Alexander Meiklejohn.¹⁹⁵ Considering the First Amendment within the historical context of its adoption and taking into account “the intention and structure of the Constitution as a whole,” Meiklejohn asserts that the main purpose of this amendment is to preserve the self-governing powers of the people.¹⁹⁶ According to this view, the First Amendment “protects the freedom of those activities of thought and communication by which we [the people] ‘govern.’ It is concerned, not with a private right, but with a public power, a governmental responsibility.”¹⁹⁷ As Meiklejohn explains, “[i]n the specific language of the Constitution, the governing activities of the people appear only in terms of casting a ballot”; however, “in the deeper meaning of the Constitution, voting is merely the external expression of a wide and diverse number of activities by means of which citizens attempt to meet the responsibilities... which [the] freedom to govern lays upon them.”¹⁹⁸ According to Meiklejohn, these responsibilities include, first, understanding the challenges that face the nation; second, passing judgment upon the decisions that the government makes to address those challenges; and third, devising methods whereby such decisions can become wiser and more effective.¹⁹⁹

¹⁹² The text of the First Amendment only refers to the federal government. *See* U.S. Const. amend. I. However, this amendment has been applied to states through the Fourteenth Amendment’s Due Process Clause. *See, e.g.*, *Gitlow v. New York*, 268 U. S. 652 (1925); *Near v. Minnesota*, 283 U.S. 697 (1931); *Grosjean v. American Press Co.*, 297 U.S. 233 (1936); *De Jonge v. Oregon*, 299 U.S. 353 (1937); *Lovell v. City of Griffin*, 303 U.S. 444 (1937); *Schneider v. State*, 308 U. S. 147 (1939), *Bridges v. California*, 314 U. S. 252 (1941).

¹⁹³ *See Sunstein, Beyond the Republican Revival*, *supra* note 30, at 1551.

¹⁹⁴ *See, e.g.*, Michelman, *Traces of Self-Government*, *supra* note 184, at 54 (suggesting that the First Amendment can be understood to envision “actual” political participation, whose “main arena is not the formal legislative assembly, but rather a dispersed and continuous process of political discussion among coconstituents [sic], and between them and their representatives.”) *See also id.* at 57.

¹⁹⁵ *See generally* ALEXANDER MEIKLEJOHN, *FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT* (1948) (arguing that there are two different types of freedom of speech); Alexander Meiklejohn, *The First Amendment Is an Absolute*, 1961 SUP. CT. REV. 245 (1961) (arguing that the First Amendment is not absolute).

¹⁹⁶ *See* Meiklejohn, *supra* note 195, at 253-254.

¹⁹⁷ *Id.* at 255. This reading of the First Amendment may be contrasted with readings that emphasize its contribution to individual liberty and autonomy. *See, e.g.*, C. Edwin Baker, *The Process of Change and the Liberty Theory of the First Amendment*, 55 S. CAL. L. REV. 293 (1982).

¹⁹⁸ *See* Meiklejohn, *supra* note 195, at 255.

¹⁹⁹ *Id.*

Meiklejohn's assertion that the main purpose of the First Amendment is to secure public engagement in governance has been endorsed by the Supreme Court in several cases, most famously in *New York Times Co. v. Sullivan*.²⁰⁰ In this case, the Court reversed a decision to award damages to a public officer following the publication of a defamatory advertisement in the New York Times, which criticized his official conduct. The Court asserted that in view of the political nature of the defamatory publication, the freedom of speech and the press must be granted particularly strong protection.²⁰¹ In reaching this conclusion, the Court invoked Madison's statement that "the Constitution created a form of government under which [t]he people . . . possess the absolute sovereignty,"²⁰² and that in such a system "the censorial power is in the people over the Government, and not in the Government over the people."²⁰³ These statements, according to the Court, revealed that the "central meaning of the First Amendment"²⁰⁴ was to provide "ample opportunity for the people to determine and resolve public issues."²⁰⁵ In subsequent First Amendment cases, the Court reaffirmed the *New York Times* principle that public speech (as opposed to private or commercial speech) deserves particularly robust—in fact, almost absolute—protection, because such speech "is more than self-expression; it is the essence of self-government."²⁰⁶

Taking the proposition that the main purpose of the First Amendment is to enable authentic public deliberation and thereby promote meaningful political participation a step further, it could be argued that this Amendment entails not only that the government should refrain from imposing any constraints upon political speech, but also that, in some cases, it should take active measures to ensure that free political speech actually takes place. The idea that a "Madisonian" reading of the First Amendment may call for its interpretation as incorporating both negative and positive elements is not new. About fifty years ago, in the heyday of "traditional" mass media, Jerome Barron famously argued that the First Amendment places upon the government a duty to ensure that private broadcasting and newspaper companies offer a stage for a variety of competing views and ideas.²⁰⁷ In recent years it has been suggested that similar requirements should be

²⁰⁰ 376 U.S. 254 (1964). For a comprehensive discussion of the relationship between Meiklejohn's thesis and the *New York Times* case, see William J. Brennan, Jr., *The Supreme Court and the Meiklejohn Interpretation of the First Amendment*, 79 HARV. L. REV. 1 (1965).

²⁰¹ The Court endorses here the concurring opinion of Justice Brandeis in *Whitney v. California*, 274 U. S. 357 (1927).

²⁰² 376 U.S. at 274.

²⁰³ *Id.* at 275.

²⁰⁴ *Id.* at 273.

²⁰⁵ *Id.* at 302 (citing WILLIAM DOUGLAS, THE RIGHT OF THE PEOPLE 41 (1st ed. 1958)).

²⁰⁶ See *Garrison v. Louisiana*, 379 U.S. 64, 74-75 (1964).

²⁰⁷ See Jerome A. Barron, *Access to the Press—A New First Amendment Right*, 80 HARV. L. REV. 1641 (1967). The Supreme Court adopted a similar approach in the *Red Lion Broadcasting* case when it upheld the right of the government to regulate broadcasting rights such that license holders would have to devote some airtime to presenting controversial issues of public importance from a

applied to the new media giants, subjecting them to neutrality principles and limiting their ability to exclusively control any particular segment of internet infrastructure or services.²⁰⁸

So far, the idea that the First Amendment may place a positive obligation upon the government to facilitate political speech has only been examined in the context of the government's alleged responsibility to regulate the activity of media companies. However, there seems to be no reason why some positive obligations concerning the freedom of speech should not apply directly to the government. According to this view, in order to realize the deliberative goals of the First Amendment, federal and state agencies should be required to offer adequate opportunities to affected citizens to express their opinions with respect to certain policy issues.

Of course, such requirements already apply to government agencies in some cases, most notably when they consider adopting new rules or repealing existing ones. In these situations, the Administrative Procedure Act requires federal agencies to publish a notice of the proposed rule and to give interested persons an opportunity to submit written comments.²⁰⁹ Similar requirements can be found in many state laws. However, according to the common understanding, these notice and comment requirements are not mandated by the Constitution, and there is no constitutional obligation upon federal or state agencies to pursue similar procedures when they make decisions that do not involve rulemaking.

By contrast, if the First Amendment is understood to place positive obligations upon government agencies, then providing opportunities to the public to participate in policymaking processes in a meaningful, deliberative way may be a matter of constitutional duty. A full examination of the possible implications of such a duty for different types of policymaking by different agencies and of the specific requirements that it might entail in the age of digital data is beyond the scope of this Article. In the context of the present study, suffice it to say that this interpretation of the First Amendment seems to present two options for an agency that uses big data analysis to guide its decisions on controversial matters of significant public interest.

The first option is to attempt to employ big data analysis in a manner that creates opportunities for the public to express diverse opinions and engage in meaningful discussion concerning the relevant policy issues. The employment of big data

variety of perspectives. However, the Court stopped short of asserting that the government had a *duty* to apply such regulations. See *Red Lion Broadcasting Co. v. Federal Communications Commission*, 395 U.S. 367 (1969).

²⁰⁸ See, e.g., Neil Weinstock Netanel, *New Media in Old Bottles? Barron's Contextual First Amendment and Copyright in the Digital Age*, 76 GEO. WASH. L. REV. 101 (2008); Hannibal Travis, *Of Blogs, eBooks, and Broadband: Access to Digital Media as a First Amendment Right*, 35 HOFSTRA L. REV. 1519 (2007) (arguing that antitrust and intellectual property law should be applied to the new media in a manner that promotes free speech).

²⁰⁹ APA, 5 U.S.C. § 553(b)-(c).

analysis in such a deliberative manner depends, among other things, on the availability of advanced machine learning algorithms that can process complex natural language texts. Such algorithms may allow for the automatic analysis of people's comments on public affairs.²¹⁰ Arguably, a governmental commitment to facilitating broad yet deliberative public involvement in policymaking—especially if backed by a legal obligation—can enhance the development of such technological capacities.²¹¹ This does not mean that once such technologies are developed, government agencies should only use big data analysis in a deliberative manner. They may well rely on technical data that people produce passively or automatically to inform their policies. However, they may not rely exclusively on such technical data.

The second option is to complement big data analysis by other methods of public participation that allow for meaningful deliberation, such as face-to-face or online consultations with those who have a special interest in the policy issues at stake. Such parallel employment of quantitative and qualitative participation methods can satisfy free speech requirements at least until such time as big data analysis technologies allow for a better combination of the two.

B. Exclusion of Disadvantaged Groups and the Fourteenth Amendment

Reading the First Amendment as imposing positive duties upon the government may also bear implications for the problem of exclusion from big data. This reading suggests that, to the extent that the government uses big data as a platform for public deliberation, it must ensure that all those affected have access to this platform. However, as implied above, even if at some point government agencies begin to use big data in a deliberative manner, relying on advanced machine learning technologies, they will probably do so only with respect to contested policy issues that involve considerable human discretion. In other cases, they are likely to use big data in a technical, non-deliberative manner, taking advantage of the massive amount of digital data that is already out there. The contribution of technical data to governmental analysis is thus likely to become and remain a major way of influencing public policy, which should be open to everyone. The First Amendment, of course, cannot be invoked to establish a right to participate in non-

²¹⁰ See, e.g., WP BrandStudio, *Future-proof: How Today's Artificial Intelligence Solutions are Taking Government Services to the Next Frontier*, WASH. POST (22 Aug. 2017). Cited in Benvenisti, *supra* note 14.

²¹¹ For example, the Department of Health and Human Services has recently tested a machine learning tool that can automatically process public comments on proposed regulations. The purpose of the project was to increase the efficiency of informal rulemaking procedures mandated by the Administrative Procedure Act (see *supra* note 209). See OFF. OF THE CHIEF TECH. OFFICER, INCREASING EFFICIENCY IN RULE MAKING WITH NATURAL LANGUAGE PROCESSING, <https://www.hhs.gov/idealab/projects-item/increasing-efficiency-in-rule-making-with-natural-language-processing> (stating that HHS Idea Labs project aimed to increase the efficiency of processing public comments). Cited in Benvenisti, *supra* note 14.

deliberative big data analysis, for the provision of blood test results or navigation data to the government has nothing to do with free speech. Instead, the most relevant source for asserting a right to be included in governmental big data analysis seems to be the Equal Protection Clause, which applies to state governments by virtue of the Fourteenth Amendment, and to the federal government by virtue of the Fifth Amendment.²¹²

According to its prevailing interpretation, the Equal Protection Clause prohibits governmental authorities from purposefully treating similarly situated persons differently.²¹³ A court is most likely to find that this prohibition has been violated when it employs a strict or heightened level of scrutiny, which is the case when the group allegedly discriminated against is considered a protected class.²¹⁴ Of course, the decision of the court as to whether a violation has occurred depends on various factors, including the existence of a public interest that could justify disparate treatment.²¹⁵ However, even without delving into these factors, it seems safe to assume that under the prevailing interpretation of the Equal Protection Clause, exclusion from big data is unlikely to be deemed unconstitutional, for two main reasons.

First, antidiscrimination jurisprudence has been reluctant to recognize socioeconomically disadvantaged people as members of a protected or quasi-protected class.²¹⁶ Although socioeconomic marginalization in general, and

²¹² The Fourteenth Amendment provides that “[n]o State shall . . . deny to any person within its jurisdiction the equal protection of the laws.” U.S. CONST. amend. XIV, § 1. This Equal Protection Clause only applies to state and local governments. However, the United States Supreme Court has interpreted the Equal Protection Clause of the Fourteenth Amendment to apply to the federal government through the Fifth Amendment’s Due Process Clause. U.S. CONST. amend. V. See *Bolling v. Sharpe*, 347 U.S. 497, 499–500 (1954) (holding that discrimination by the federal government may be violative of the Fifth Amendment’s Due Process Clause).

²¹³ See *Washington v. Davis*, 426 U.S. 229, 239–242 (1976) (holding that in order to establish a violation of the Equal Protection Clause, a claimant must show that discrimination was intentional, and noting that a disparate outcome in itself does not automatically prove discrimination). See also *Akins v. Texas*, 325 U.S. 398, 403–404 (emphasizing that discrimination must be intentional).

²¹⁴ See, e.g., *Romer v. Evans*, 517 U.S. 620 (1996) (“[I]f a law neither burdens a fundamental right nor targets a suspect class, we will uphold the legislative classification so long as it bears a rational relation to some legitimate end.”). There is, however, no clear definition of what constitutes a protected class. See, e.g., Marcy Strauss, *Reevaluating Suspect Classifications*, 35 SEATTLE U. L. REV. 135, 138 (2011) (“The Supreme Court has not provided a coherent explanation for precisely what factors trigger heightened scrutiny”); Reva B. Siegel, *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles Over Brown*, 117 HARV. L. REV. 1470, 1542 (2004) (“American antidiscrimination law has no determinate criteria for deciding what practices are group-based classifications, and while courts sometimes articulate such criteria, they often apply them inconsistently”).

²¹⁵ See, e.g., *City of Cleburne v. Cleburne Living Ctr.*, 473 U.S. 432, 439–440 (1985).

²¹⁶ See, e.g., *Harris v. McRae*, 448 U.S. 297, 323 (1980) (noting that the Supreme Court has repeatedly held that poverty is not a suspect classification); *Maher v. Roe*, 432 U.S. 464, 471 (1977) (asserting that financial need alone does not identify a suspect class); *San Antonio Indep. Sch. Dist. v. Rodriguez*, 411 U.S. 1, 28 (1973) (explaining that people living in poor districts are not a protected

exclusion from big data in particular, often coincides with characteristics that otherwise establish protected classes, such as race, ethnicity, and gender,²¹⁷ these characteristics cannot be said to distinguish big data outsiders as a group. Instead, big data outsiders represent a “large, diverse, and amorphous class” to which equal protection guarantees are rarely applied.²¹⁸

Second, in most cases, it would be hard to infer a discriminatory intent from the fact that socioeconomically disadvantaged people contribute to big data less than others. As we have seen, governmental authorities often harvest big data from a wide range of sources over which they do not have full control. It makes little sense to claim that the health authorities choose to collect medical data from EHRs or social sites because poor people do not have adequate access to these sources, or that urban planners collect data from mobile applications precisely for this reason. Even when the authorities are involved in the unequal distribution or installation of sensors or other devices from which they hope to collect big data, they do not necessarily do so with the intention to discriminate. Indeed, the problem with big data-based policies is not that they intentionally ignore the needs of poorer or older populations, but rather that they fail to acknowledge the economic, socio-psychological, and cultural constraints that prevent these populations from producing big data in the same amount as better-off groups.

It seems, then, that the existing equal protection doctrine, which requires an intentional disparate *treatment* of protected groups in order to establish unlawful discrimination, cannot pave the way for the inclusion of socioeconomically disadvantaged populations in big data. As is well known, the disparate treatment doctrine has been criticized by many academics and civil rights advocates who have suggested that the Equal Protection Clause should be understood to prohibit not only intentional disparate treatment but also unintentional disparate *impact*.²¹⁹

class because they are “not saddled with such disabilities, or subjected to such a history of purposeful unequal treatment, or relegated to such a position of political powerlessness as to command extraordinary protection from the majoritarian political process.”). Cf. Henry Rose, *The Poor as a Suspect Class under the Equal Protection Clause: An Open Constitutional Question*, 34 NOVA L. REV. 407 (2010) (arguing that the Supreme Court’s decision in *San Antonio Indep. Sch. Dist. v. Rodriguez* has been misunderstood and that, in fact, the Court has not yet decided whether the poor are a quasi-suspect or suspect class). The term “quasi-protected class” refers to “a class of individuals who lack political power and have been subjected to purposeful, intentional discrimination to a lesser degree” than individuals in a fully protected class. See Jennifer E. Watson, *When No Place Is Home: Why the Homeless Deserve Suspect Classification*, 88 IOWA L. REV. 501, 508 (2003).

²¹⁷ See *supra* note 36 and accompanying text.

²¹⁸ *San Antonio Indep. Sch. Dist.*, 411 U.S. at 28. See also Goodwin Liu, *Education, Equality, and National Citizenship*, 116 YALE L.J. 330, 334 (2006) (noting that “equal protection has been less potent in addressing disadvantage... that affects a diffuse or amorphous class”); Lerman, *supra* note 54, at 60 (citing Liu).

²¹⁹ See, e.g., Reva Siegel, *Why Equal Protection No Longer Protects: The Evolving Forms of Status-Enforcing State Action*, 49 STAN. L. REV. 1111, 1131 (1997) (arguing that the Supreme Court’s current interpretation of the Equal Protection Clause authorizes state action that perpetuates historic

According to this approach, governmental practices that are neutral on their face but have a disproportionately adverse impact on protected classes should also be considered unconstitutional.²²⁰ Unlike the disparate treatment standard, which focuses on the formal classifications underlying particular governmental decisions, disparate impact is concerned with the actual effects of the ongoing subordination of historically disempowered groups.²²¹ Had this standard been adopted by the Supreme Court, big data-based governmental policies that disfavor socioeconomically disadvantaged populations might have been deemed unconstitutional.²²² For although big data is perfectly neutral on its face and seems to provide an Arendtian mask that denies discriminatory classifications,²²³ it can actually perpetuate the underrepresentation and political subordination of historically disadvantaged groups.

Another somewhat speculative constitutional solution to the problem of big data exclusion may be found in the Privileges or Immunities Clause of the Fourteenth Amendment, which precedes the Equal Protection Clause.²²⁴ This Clause, which provides that states shall not abridge the privileges or immunities of the citizens of the United States, grants some entitlements to all citizens of the United States and asserts that states must respect these entitlements.²²⁵ However, the content and scope of these entitlements are not explicitly mentioned in the Constitution and have been the subject of much controversy.²²⁶ In the 1872 *Slaughter-House Cases*,

forms of race and gender stratification); Barbara J. Flagg, “*Was Blind, But Now I See*”: *White Race Consciousness and the Requirement of Discriminatory Intent*, 91 MICH. L. REV. 953, 968 (1993) (contending that “the Davis rule reflects a distinctively white way of thinking about race”); Charles R. Lawrence, III, *The Id, the Ego, and Equal Protection: Reckoning With Unconscious Racism*, 39 STAN. L. REV. 317, 323 (1987) (asserting that “requiring proof of conscious or intentional motivation as a prerequisite to constitutional recognition that a decision is race-dependent ignores much of what we understand about how the human mind works”); Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 MINN. L. REV. 1049, 1052 (1978) (criticizing the Supreme Court’s focus on discriminatory intent).

²²⁰ See, e.g., Michael Selmi, *Was the Disparate Impact Theory a Mistake?* 53 UCLA L. Rev. 701, 708 (2006) (observing that disparate impact theorists are concerned with the perpetuation of past lawful discrimination through what appear to be neutral practices).

²²¹ See, e.g., Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 PHIL. & PUB. AFF. 107, 108 (1976) (arguing that the approach that focuses on formal classifications is highly individualistic and does not take social reality fully into account).

²²² See Lerman, *supra* note 54, at 57, 61-62 (arguing that a new big data antisubordination doctrine may be needed to protect the persons whom the big data revolution risks sidelining).

²²³ See *supra* notes 103-110, 112-113, and accompanying text.

²²⁴ “No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States.” U.S. CONST. amend. XIV, § 1.

²²⁵ *Id.*

²²⁶ See, e.g., Emily Jennings, Let’s All Agree to Disagree, and Move On: Analyzing Slaughter-House and the Fourteenth Amendment’s Privileges or Immunities Clause Under “Sunk Cost” Principles, 54 BOSTON COLLEGE L. REV. 1803, 1806-1815 (2013) (detailing the disagreements among legal historians with respect to the intention of the drafters of the Privileges or Immunities Clause); JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 98 (3rd ed.

the Supreme Court interpreted the Privileges or Immunities Clause very narrowly, rejecting the view that it protects all the basic rights of U.S. citizens against infringement by the states, and instead holding that it secures only a limited number of rights that “owe their existence to the federal government,” such as the right to interstate travel and the right to use the nation’s navigable waters.²²⁷ According to the Court’s reading, the Clause preserves rather than limits state authority over civil rights, subjecting it only to “an anemic and eclectic array”²²⁸ of external rights that stem directly from federal citizenship. Following the *Slaughter-House* decision, which was reiterated by the Supreme Court in subsequent cases,²²⁹ the Privileges or Immunities Clause has lain nearly dormant for 145 years.²³⁰

However, although *Slaughter-House* remains on the books, many legal scholars maintain, along with the minority judges in *Slaughter-House*, that this decision represents a flawed interpretation of the Privileges or Immunities Clause.²³¹ According to these critics, the Clause was meant to protect against state abridgement of a broader list of substantive rights than *Slaughter-House* recognized.²³² More specifically, some critics suggest that, read together with the

1981) (noting that the Privileges or Immunities Clause “is quite inscrutable, indicating only that there should exist some set of constitutional entitlements not explicitly enumerated in the document”).

²²⁷ See *Butchers’ Benevolent Assoc. of New Orleans v. Crescent City Live-Stock Landing & Slaughter-House Co.*, 83 U.S. (16 Wall.) 36, 74-75, 78-80 (1872) [hereinafter *Slaughter-House*].

²²⁸ Goodwin Liu, *supra* note 218, at 354.

²²⁹ In the few cases since *Slaughter-House* where petitioners attempted to expand the scope of the Privileges or Immunities Clause, the Supreme Court refused to revive it. See, e.g., *United States v. Cruikshank*, 92 U.S. 542, 551– 53 (1875) (holding that the Privileges or Immunities Clause did not protect the right to peaceably assemble to petition state policies); *McDonald v. City of Chicago*, 130 S. Ct. 3020 (2010) (refusing to assert that the Privileges or Immunities Clause required states to respect the right to bear arms, and preferring to apply this duty to states through the Fourteenth Amendment’s Due Process Clause).

²³⁰ As Alexander M. Bickel has observed, the *Slaughter-House* decision “just about read the privileges and immunities clause out of the Constitution.” See Alexander M. Bickel, *Citizenship in the American Constitution*, 15 ARIZ. L. REV. 369, 378 (1973).

²³¹ See *Slaughter-House*, 83 U.S. (16 Wall.) at 83, 96-101 (Field, J., dissenting) (maintaining that the majority’s interpretation of the Clause made it a “vain and idle enactment, which accomplished nothing,” and suggesting instead that it was meant to expand the protections provided by the Comity Clause of Article IV, Section 2 of the Constitution); *id.* at 111, 118-119 (Bradley, J., dissenting) (opining that the Privileges or Immunities Clause was intended to secure the federal rights enumerated in the Constitution’s Bill of Rights against state infringement). See generally Richard L. Aynes, *Ink Blot or Not: The Meaning of Privileges and/or Immunities*, 11 U. PA. J. CONST. L. 1295, 1295-96 (2009) (arguing that the Privileges or Immunities Clause incorporates the Bill of Rights); Kimberly C. Shankman & Roger Pilon, *Reviving the Privileges or Immunities Clause to Redress the Balance Among States, Individuals, and the Federal Government*, 3 TEX. REV. L & POL. 1 (1998) (same); Akhil Reed Amar, *The Bill of Rights and the Fourteenth Amendment*, 101 YALE L.J. 1193 (1992) (same).

²³² See, e.g., Rebecca E. Zietlow, Belonging, Protection and Equality: The Neglected Citizenship Clause and the Limits of Federalism, 62 U. PITTS. L. REV. 281, 313-14 (2000) [hereinafter Zietlow, Belonging, Protection and Equality] (asserting that there is considerable evidence that the Framers

Citizenship Clause of the Fourteenth Amendment,²³³ the Privileges or Immunities Clause should be understood to guarantee the equal right of all citizens of the United States to enjoy the benefits of national citizenship regardless of their state of residence.²³⁴ According to Goodwin Liu, this guarantee of equal national citizenship places a duty upon both state governments and the federal government to secure for all citizens “full membership [and] effective participation.... in the national community.”²³⁵ Liu and other legal historians find evidence in the debates that surrounded the ratification of the Fourteenth Amendment and in contemporaneous legal doctrine and commentary that the Citizenship Clause and the Privileges or Immunities Clause were intended to “encompass substantive rights necessary to make citizenship meaningful and effective.”²³⁶ The Framers understood, however, that “citizenship was an evolving concept” and therefore chose to employ the “broad language of ‘Privileges or Immunities’” that would “enable future generations...to develop further the privileges and immunities of citizenship.”²³⁷

If we accept the propositions that the Privileges or Immunities Clause is intended to secure equal citizenship and effective participation in political life and that the exact meaning of these guarantees should be determined in accordance with the realities of a given period, we can argue that in the second decade of the twenty-first century, this Clause protects the right of all citizens to equal representation in big data that shapes public policy.²³⁸ We can also argue that this right is enforceable against both state (and local) governments and the federal government, and that it places upon them an affirmative duty to promote big data equality.²³⁹ Given the fact

intended the Privileges or Immunities to apply very broadly and to “encompass all fundamental human rights and link those rights to national citizenship”).

²³³ “All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside.” U.S. CONST. amend. XIV, § 1.

²³⁴ See, e.g., Liu, *supra* note 218; Zietlow, *Belonging, Protection and Equality*, *supra* note 232, at 307-333.

²³⁵ Liu, *supra* note 218, at 335

²³⁶ Liu, *supra* note 218, at 357. See also Rebecca E. Zietlow, *Congressional Enforcement of Civil Rights and John Bingham’s Theory of Citizenship*, 36 AKRON L. REV. 717, 739 (2003) (“Evidence from the Ratification debates and contemporaneous legal doctrine indicates that the Framers viewed the meaning of federal citizenship very broadly and that the rights that adhered to citizenship were considerably broader than those enumerated in the Bill of Rights.”).

²³⁷ James W. Fox, Jr., *Citizenship, Poverty, and Federalism: 1787-1882*, 60 U. PITTS. L. REV. 421, 504 (1999). See also Daniel A. Farber & John E. Muench, *The Ideological Origins of the Fourteenth Amendment*, 1 CONST. COMMENT. 235, 274-275 (1994) (“[T]he Framers realized... that the amendment’s precise content in some sense was in the hands of future generations.”); Robert J. Kaczorowski, *Revolutionary Constitutionalism in the Era of the Civil War and Reconstruction*, 61 N.Y.U. L. REV. 863, 923-926 (1986).

²³⁸ Goodwin Liu reaches a similar conclusion with respect to equal opportunities in education, arguing that the Fourteenth Amendment authorizes and obligates Congress to ensure a meaningful floor of educational opportunity throughout the nation. See Liu, *supra* note 218.

²³⁹ Section 5 of the Fourteenth Amendment provides that “[t]he Congress shall have power to

that the Privileges or Immunities Clause has drawn far less judicial attention than the Equal Protection Clause, it may be easier for courts and governmental authorities to adopt a broad interpretation of the former that encompasses big data equality than to admit the illegality of big data's disparate impact under the latter.

Assuming that the Fourteenth Amendment—by virtue of either the Equal Protection Clause or the Privileges or Immunities Clause—requires that big data-based policymaking be inclusive of all segments of the population, what should state and federal authorities do to meet this requirement? At the least, it seems, they need to address the two basic dimensions of the digital divide, namely, physical access and mental skills.²⁴⁰ This means, first, that government authorities should make every effort to ensure that the digital sources from which they collect big data are accessible to everyone. For example, a local authority that relies on data produced by personal smart-city sensor kits to address noise and air pollution could offer such sensors for free to those who cannot afford to purchase them. Or, if the authority collects data from stationary sensors that it installs in public places, it should make sure that the sensors are distributed evenly across town. Similarly, if health authorities wish to harvest medical data from personal health monitors, they should make sure that the cost of the latter does not prevent potential participants from using them. Of course, in some cases it may be more complicated to expand access to digital sources. For example, to ensure equal representation in EHRs may take a fundamental reform in federal health insurance coverage, which does not seem feasible at this time.²⁴¹ It may be possible, however, to find alternative ways to collect data relating to the medical needs of uninsured persons and to take it into account when shaping health policies. Second, government authorities should develop and implement programs to facilitate diverse uses of digital platforms by all citizens. These programs should help people acquire the technical skills that allow them to produce digital data, and they should also raise awareness of the potential policy implications of producing such data, so as to allow people to make informed choices about their digital activities. In addition to taking active measures to promote inclusiveness, government authorities should make sure to abide by transparency principles that allow those affected by big data-based policymaking to know, to the extent possible, which data was used, how it was collected, and from

enforce, by appropriate legislation, the provisions of this article.” U.S. CONST. amend. XIV, § 5. Many commentators assert that the enforcement power granted to the Congress in Section 5 also places upon it *an obligation* to enforce the Fourteenth Amendment, including the Privileges or Immunities Clause. See, e.g., Liu, *supra* note 218; Zietlow, *Belonging, Protection and Equality*, *supra* note 232, at 310 (“[T]he Citizenship Clause also invokes the power of the federal government to protect the rights that adhere to federal citizenship.”); Farber & Muench, *supra* note 237, at 236 (“The fourteenth amendment was intended to bridge the gap between positive law and higher law by empowering the national government to protect the natural rights of its citizens.”); William J. Rich, *Taking “Privileges or Immunities” Seriously: A Call to Expand the Constitutional Canon*, 87 MINN. L. REV. 153 (2002) (arguing that Congress should enforce the privileges or immunities of citizenship).

²⁴⁰ See *supra* notes 36-40 and accompanying text.

²⁴¹ See *supra* notes 137-139 and accompanying text.

whom.

As noted above, much of the digital data that government authorities employ to inform policymaking is originally collected by private entities (e.g., medical clinics, health insurers, internet giants, and mobile software developers) for their own non-policy-oriented purposes. Hence, in order to ensure that the data that they analyze is inclusive, government authorities may have to place some inclusiveness requirements upon those private data providers. These entities can be required to facilitate access to certain services that they provide, which are of potential use to policymakers. They can also be required to actively encourage disengaged populations to use digital technology in a manner that has the potential to affect public policy, for example by soliciting their inputs on public affairs or by nudging them to produce other policy-relevant information. Finally, private data collectors should also be subjected to transparency requirements that enable public scrutiny of non-policy-oriented data used by the government. These requirements may be softer than the ones that apply directly to the government and may be subject to trade secret protection and other exceptions. While all these regulatory measures, even if moderate, are likely to meet with some resistance on the part of powerful private data collectors, the experience of recent years shows that governmental attempts to regulate the big data industry can nonetheless be effective. So far, such regulation has mostly been designed to protect customer privacy and security,²⁴² but there seems to be no reason why regulatory efforts to promote the vital interest of (and, arguably, the constitutional right to) data inclusiveness should not be equally successful.

CONCLUSION

The incorporation of big data analysis into public policymaking carries a promise not only of enhanced accuracy and efficiency, but also of greater political participation. In an ideal world, big data could allow all affected citizens to have their voice heard in policymaking processes. In the real world, however, big data fails to fulfill this potential. First, instead of providing a platform through which people can convey their reasoned opinions about contested policy questions, big data serves merely as a means for the government to aggregate and analyze

²⁴² See, e.g., Federal Communications Commission, Protecting the Privacy of Customers of Broadband and Other Telecommunications Services, Report and Order (Broadband Privacy Order), DA/FCC: FCC-16-148 (Oct. 27, 2016) (empowering broadband consumers to decide how data is used and shared by broadband providers); Fair and Accurate Credit Transactions Act of 2003, 15 U.S.C. § 1681 (2012) (allowing consumers to access their credit reports and prescribing measures to reduce the risk of identity theft); Federal Information Security Management Act of 2002 (FISMA), 4 U.S.C. §§3541-3549 (2006) (creating a security framework for federal information systems); Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501-6506 (2012) (regulating the online collection of personal information from children under the age of thirteen); Health Insurance Portability and Accountability Act (HIPAA) of 1996, 42. U.S.C. § 1320d-2 (2012) (requiring healthcare entities to secure the confidentiality of personal health information).

technical data about people's habits and conditions. Second, instead of functioning as a mask that covers physical characteristics such as gender and race and thus allows people to participate in collective decision-making on an equal basis, the apparent neutrality of big data allows for the exclusion of disadvantaged populations to go unnoticed, thereby reinforcing political inequalities.

Constitutional law can offer a normative framework for addressing these deficits and reducing the gap between the promise and reality of big data-based policymaking. This Article has presented a somewhat uncommon yet plausible interpretation of the First and Fourteenth Amendments that may establish a constitutional right to meaningful and inclusive participation in governmental big data analysis. This interpretation takes into account the original intention of the Constitution's Framers and the values that they sought to promote, but at the same time it acknowledges the need to adapt these values to contemporary political and technological circumstances. Although this interpretation does not find full support in existing judicial decisions, it may be adopted in the future when courts will have to face the challenges that rapidly emerging big data applications pose to legal doctrine created in the pre-big data era. If this happens, government authorities will have to find ways to make big data-based policymaking more deliberative and inclusive. The Article has proposed possible directions towards accomplishing this task, which need to be further examined and refined against real-life technological and political developments.